Methods in Molecular Biology 1334

Springer Protocols

Benoît P. Leblanc
Sébastien Rodrigue *Editors*

DNA-Protein Interactions

Principles and Protocols

Fourth Edition

Humana Press

METHODS IN MOLECULAR BIOLOGY

Series Editor
John M. Walker
School of Life and Medical Sciences
University of Hertfordshire
Hatfield, Hertfordshire, AL10 9AB, UK

For further volumes:
http://www.springer.com/series/7651

DNA-Protein Interactions

Principles and Protocols

Fourth Edition

Edited by

Benoît P. Leblanc and Sébastien Rodrigue

Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada

Humana Press

Editors
Benoît P. Leblanc
Département de Biologie, Faculté des Sciences
Université de Sherbrooke
Sherbrooke, QC, Canada

Sébastien Rodrigue
Département de Biologie, Faculté des Sciences
Université de Sherbrooke
Sherbrooke, QC, Canada

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-4939-2876-7 ISBN 978-1-4939-2877-4 (eBook)
DOI 10.1007/978-1-4939-2877-4

Library of Congress Control Number: 2015945654

Springer New York Heidelberg Dordrecht London

Printed on acid-free paper

Humana Press is a brand of Springer
Springer Science+Business Media LLC New York is part of Springer Science+Business Media (www.springer.com)

Preface

Science and technology have gone hand in hand for several centuries now, to the point where they are essentially synonymous in the mind of the general public. Considering the immense progress made since the Enlightenment in the fields of chemistry, physics, and biology (to name only those disciplines) due to the availability of ever more sophisticated measuring devices, we must admit, if a bit ruefully, that our understanding of the universe would have been a very limited thing indeed if all we had to rely on had been our five senses and our intelligence. And so a curious and industrious humanity built tools that would extend the reach of its limited senses to bring previously unreachable knowledge within its intellectual grasp. New technologies thus help broaden our basic knowledge, which in turn makes even more advanced tools possible, an incremental process that shows no sign of slowing down. And once in a while, a particular technological development comes around and has such a tremendous impact that we never do things the same way again.

Molecular biology has known several of these world-changing technical revolutions during the short few decades of its existence as a discipline. The most famous of these is probably the polymerase chain reaction (PCR), which opened whole new vistas of gene recovery and targeted mutations. Another field-altering technique, which came a little later, is chromatin immunoprecipitation (mostly referred to as ChIP). ChIP allowed us to work in an in vivo context (or at least ex vivo one, for the purists out there) as easily as we previously had worked with in vitro models, and got us one step closer to understanding what really goes on in living cells. This technique in turn greatly benefited from the invention of real-time, quantitative PCR (QPCR), as it then allowed us to easily quantify the degree of interaction of just about any protein with its cognate binding sites in the genome.

In the last few years, we also welcomed the development of affordable large-scale sequencing technologies, which coupled to the aforementioned techniques turned the world on its head again. The impact of massive parallel sequencing is reflected by the widespread use of the suffix "seq" added to the name of any number of today's molecular biology techniques, almost rivaling in its ubiquity the prefix "i" added to the name of today's electronic gizmos. These technologies, whether they be applied to the rapid deciphering of whole genomes or to the identification of DNA fragments such as could be isolated by chromatin immunoprecipitation, make it possible to ask questions about the genome as a whole rather that about individual genes in a way that would have been inconceivable a mere 10 years ago.

Naturally, asking questions in a genomic context forces us to leave behind the concept of DNA as a naked molecule and to ponder the mysteries of chromatin structure. Terms like "histones" and "nucleosomes," as well as "epigenetics," have become very trendy in the past decade. In fact, it is not rare to see someone arguing that all there is to know about gene expression revolves around covalent modifications of histones, or of modification of DNA itself, and that epigenomics is all about acetyltransferase or methyltransferase enzymes… making the study of the interaction between DNA sequences and transcription factors something a little *passé*, something a little quaint.

Well, if covalent modifications of chromatin components manage to play such an undeniably important role, it is certainly because the relevant enzymes were directed to the proper sites on the chromosomes… and that crucial part of gene expression regulation is still the responsibility of hundreds of activators and repressors that bring to the cocktail an important element: sequence specificity. Without these factors, it would be difficult indeed for a development program to bring about the appropriate differentiation of cells, tissues, and organs in an embryo, or for the cell to respond in an appropriate fashion to all the signals that inform it about its own condition and that of its environment. All proteins interacting with DNA play a role in the genome's *modus operandi*, and we would be remiss to neglect any of them in our quest to understand how the cell functions and how organisms manage to maintain their homeostasis.

As we head into an era of personal genomics, it will be more important than ever to assess how this or that protein interacts with hundreds and thousands of DNA control elements, so that gene-based therapeutic strategies may influence expression in a beneficial way while limiting side effects. Technically speaking, we have come a very long way since the early gel shifts, but the DNA-protein interaction game is still about what sticks where, and how strongly.

This fourth edition of *DNA-Protein Interaction: Principles and Protocols* is made necessary by the recent technical advances mentioned above. Some chapters found in previous editions, describing techniques that are still powerful despite their simplicity or have proven their continuous worth, have been updated as needed. Meanwhile, many new chapters have been added that mostly deal with larger-scale experiments, reflecting recent advances in "big biology." We believe the result is a well-rounded volume where the best of the past is combined with the best of today, and that will offer a very useful compendium of protocols allowing one to delve into the intricacies of protein-DNA interaction at levels ranging from the very small (as in the case of single-molecule FRET) to the very complex (as with circular chromatin conformation capture, or 4C). In all cases, these protocols will include my favorite thing about the *Methods in Molecular Biology* series: the notes section, where scientists allow us to look behind the curtain, as it were, and have the opportunity to explain in an almost informal way how they really do things and how to avoid the pitfalls that they had to face themselves. Because let's face it: there are many more ways to mess up an experiment than to do it right, and it is good to have the detailed advice of an expert before we attempt it.

We would like to thank all the authors who contributed protocols to this book and agreed to share their valuable expertise. We hope the karmic wheel will reward them profusely. We would also like to thank our predecessors, Professors Geoff Kneale and Tom Moss, who captained the ship for the first and second editions of *DNA-Protein Interactions.* It was a delight to work with Tom for the third edition as well. These gentlemen set a fine example indeed, and we hope the current edition will prove worthy of it.

Now on to the discovery of how life functions!

Sherbrooke, QC, Canada ***Benoît P. Leblanc***
Sébastien Rodrigue

Contents

Contributors

Virginia Babl • *Lehrstuhl Biochemie III, Biochemie-Zentrum Regensburg (BZR), Universität Regensburg, Regensburg, Germany*

Nathalie Bastien • *Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, QC, Canada*

Stéphanie Bianco • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*

Simon Blouin • *Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, QC, Canada*

Julien Boudreault • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*

Philippe Bouvet • *Laboratoire Joliot-Curie CNRS USR 3010, Ecole Normale Supérieure de Lyon, Université de Lyon, Lyon, France; Laboratoire de Biologie Moléculaire de la Cellule, CNRS UMR 5239, Lyon, France*

Mylène Brunelle • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*

Vincent Burrus • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*

Adrien Chauvier • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*

David J. Clark • *Laboratory of Molecular Growth Regulation, National Institute of Child Health and Human Development (NICHD), National Institutes of Health (NIH), Bethesda, MD, USA*

Charles Coulombe • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*

Timothy D. Craggs • *Department of Physics, Oxford University, Oxford, UK*

Régen Drouin • *Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, QC, Canada*

Mayra Furlan-Magaril • *Nuclear Dynamics Programme, The Babraham Institute, Cambridge CB22 3AT, UK; Instituto de Fisiología Celular, Departamento de Genética Molecular, Universidad Nacional Autónoma de México, Mexico City*

Luc Gaudreau • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*

Nicolas Gévry • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*

Frédéric Grenier • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*

Joachim Griesenbeck • *Lehrstuhl Biochemie III, Biochemie-Zentrum Regensburg (BZR), Universität Regensburg, Regensburg, Germany*

Deborah Hinton • *Laboratory of Cell and Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health (NIH), Bethesda, MD, USA*

Meng-Lun Hsieh • *Laboratory of Cell and Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health (NIH), Bethesda, MD, USA*
Pierre-Étienne Jacques • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada; Département d'informatique, Faculté des sciences, Université de Sherbrooke, Sherbrooke, QC, Canada; Centre de recherche du Centre Hospitalier Universitaire de Sherbrooke, Sherbrooke, QC, Canada*
Tamara James • *Laboratory of Cell and Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health (NIH), Bethesda, MD, USA*
Harry Jarrett • *Department of Chemistry, University of Texas at San Antonio, San Antonio, TX, USA*
Yinshan Jia • *Department of Chemistry, University of Texas at San Antonio, San Antonio, TX, USA*
Geoffrey G. Kneale • *Institute of Biomedical and Biomolecular Sciences, School of Biological Sciences, University of Portsmouth, Portsmouth, UK*
Leslie Knipling • *Laboratory of Cell and Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health (NIH), Bethesda, MD, USA*
Daniel A. Lafontaine • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*
Benoît P. Leblanc • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*
Jean-François Lucier • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*
Alexei Nordell Markovits • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*
Dominick Matteau • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*
Philipp Milkereit • *Lehrstuhl Biochemie III, Biochemie-Zentrum Regensburg (BZR), Universität Regensburg, Regensburg, Germany*
Jean-François Millau • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*
Tom Moss • *Département de Biologie Moléculaire, Biochimie Médicale et Pathologie, Université Laval, QC, Canada*
Bruce D. Murphy • *Centre de Recherche en Reproduction Animale, Faculté de Médecine Vétérinaire, Université de Montréal, Saint Hyacinthe, QC, Canada*
Linda Nagore • *Department of Chemistry, University of Texas at San Antonio, San Antonio, TX, USA*
Isabelle Paradis • *Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, QC, Canada*
J. Carlos Penedo • *School of Physics and Astronomy, University of St. Andrews, St. Andrews, UK*
D. Cibran Perez-Gonzalez • *School of Physics and Astronomy, University of St. Andrews, St. Andrews, UK*
Christian Poitras • *Institut de Recherches Cliniques de Montréal, Montréal, QC, Canada*

Gregory M.K. Poon • *Department of Chemistry, Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA, USA*
Dominic Poulin-Laprade • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*
Félix Recillas-Targa • *Instituto de Fisiología Celular, Departamento de Genética Molecular, Universidad Nacional Autónoma de México, Mexico City*
François Robert • *Institut de Recherches Cliniques de Montréal, Montréal, QC, Canada;Département de Médecine, Faculté de Médecine, Université de Montréal, Montréal, QC, Canada*
Marc-Antoine Robert • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*
Sébastien Rodrigue • *Département de Biologie, Faculté des Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada*
Garry Scarlett • *Institute of Biomedical and Biomolecular Sciences, School of Biological Sciences, University of Portsmouth, Portsmouth, UK*
Giuliano Siligardi • *Diamond Light Source Ltd., Didcot, Oxfordshire, UK*
Victor Y. Stefanovsky • *Laboratory of Growth and Development, St-Patrick Research Group in Basic Oncology, Cancer Division of the Quebec University Hospital Research Centre, Québec, QC, Canada; Department of Molecular Biology, Medical Biochemistry and Pathology, Faculty of Medicine, Laval University, Québec, QC, Canada*
Ulrike Stöckl • *Lehrstuhl Biochemie III, Biochemie-Zentrum Regensburg (BZR), Universität Regensburg, Regensburg, Germany*
Herbert Tschochner • *Lehrstuhl Biochemie III, Biochemie-Zentrum Regensburg (BZR), Universität Regensburg, Regensburg, Germany*
François Vigneault • *Department of Genetics, Harvard Medical School, Boston, MA, USA*
Shuo Wang • *Department of Chemistry, Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA, USA*
W. David Wilson • *Department of Chemistry, Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA, USA*

Chapter 1

Electrophoretic Mobility Shift Assay Using Radiolabeled DNA Probes

Dominic Poulin-Laprade and Vincent Burrus

Abstract

Electrophoretic mobility shift assays (EMSA) have proven their usefulness for studying interactions between biological molecules. In the present protocol, a purified protein of interest is mixed with a 5′-end radiolabeled DNA probe. The bound complexes are separated by electrophoretic migration through a polyacrylamide gel and detected with a phosphorimager. The applications of EMSA are diverse, from thermodynamic and kinetic analyses to observation of bending and other conformational changes, stoichiometric inferences, or insights into cooperative protein binding.

Key words EMSA, Electrophoretic mobility shift assay, Gel shift, Gel retardation, DNA probe labeling, Native polyacrylamide gel, Phosphorus-32, Protein–DNA interactions

Abbreviations

APS	Ammonium persulfate
ATP	Adenosine triphosphate
Bis-Tris	1,3-bis(tris(hydroxymethyl)methylamino)propane
BSA	Bovine serum albumin
cAMP	3′-5′-cyclic adenosine monophosphate
CAP	*E. coli* cAMP receptor protein
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediamine tetraacetic acid
EMSA	Electrophoretic mobility shift assay
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
MOPS	3-(*N*-morpholino)propanesulfonic acid
PMSF	Phenylmethylsulfonyl fluoride
SELEX	Systematic evolution of ligands by exponential enrichment
SDS-PAGE	Polyacrylamide gel electrophoresis carried in presence of sodium dodecyl sulfate
TAE	Tris-acetate-EDTA
TBE	Tris-borate-EDTA

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_1,

TE	Tris-EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane

1 Introduction

1.1 The Principle

The formation of complexes between biological molecules is fundamental for life development. Studies involving association and dissociation of such complexes or the rate at which they occur are essential in the fields of drug discovery, cellular biology, DNA processing, and genetic regulation. Electrophoretic mobility shift assays have been used since the early 1980s [1] to study the DNA-binding ability of proteins, and it is still a trusted approach to assess qualitative and quantitative parameters of protein–nucleic acid interactions. The basic principle of the method is that the structure, the size, and the charge of a molecule will affect its migration through a native polyacrylamide (or agarose) gel. Typically, the negative charge of the labeled nucleic acid will drive its migration in the gel matrix up to a given position according to the conditions (gel matrix, buffer composition, duration, voltage) of the electrophoresis. Under the same conditions, binding of a protein on the DNA probe will slow down the DNA migration and shorten the distance it travels through the gel (shift) as a function of the overall charge and increased size of the protein–DNA complex.

1.2 The Binding Entities

The protocol described in this chapter focuses on 5′-end radiolabeled PCR-amplified probes bound by purified DNA-binding proteins. From this framework, derived protocols can be optimized to suit specific experimental needs and limitations. Figure 1 describes some alternatives for starting materials, detection methods, and analysis. The probe can be either single-stranded, duplex, triplex, or quadruplex DNAs or RNAs, as well as small circular DNAs [2]. Radiolabeled species can be detected at the sub-picomolar range, and nanomoles of proteins are generally sufficient to generate a shift. This high sensitivity combined with the narrow pore size of acrylamide gels, which is around 5–20 nm (50–200 Å) for 10 to 4 % acrylamide gels [3], allows the observation of subtle changes in conformation or stoichiometry of intermediate complexes. If lower sensitivity is sufficient, for example, the qualitative binding of a large protein, other labeling methods and agarose gels (700–7000 Å) can be used (*see* **Note 1**).

The methodology to produce and purify proteins of interest goes beyond the scope of the current chapter. Since they are generally soluble, DNA-binding proteins are easy to produce in most cases. For quantitative assessments of bacterial transcriptional factors, we recommend the use of recombinant strategies to purify the protein of interest by affinity chromatography. Many companies

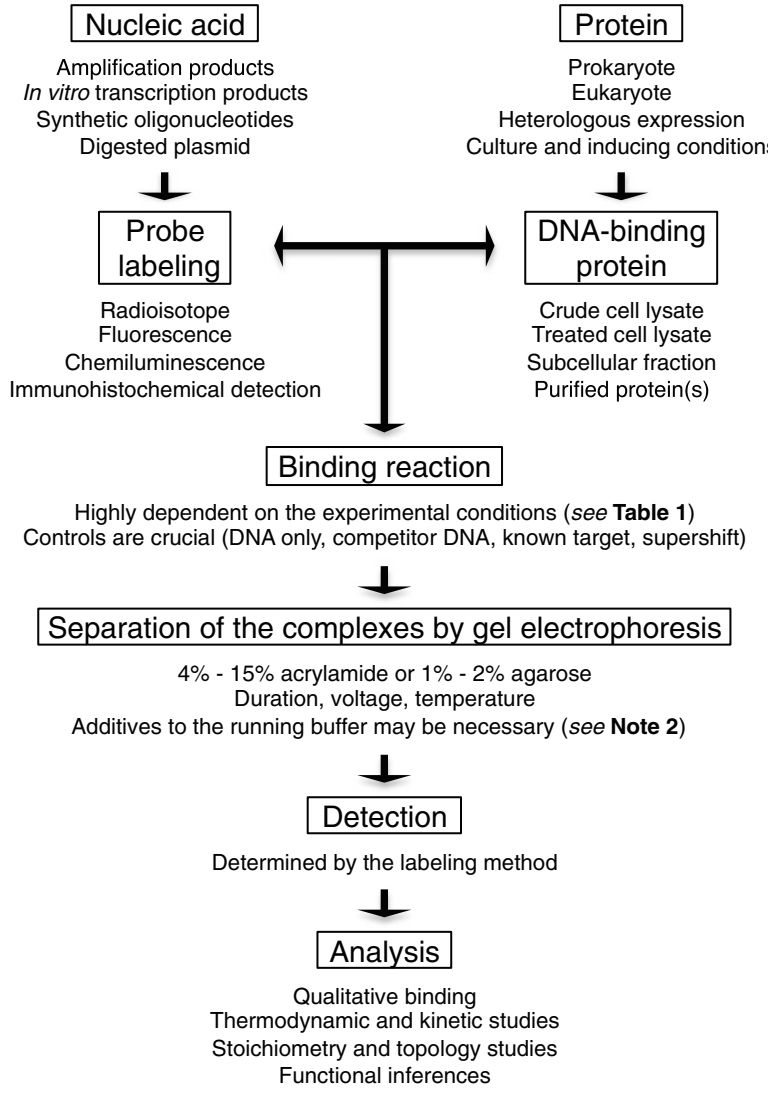


Fig. 1 Organizational chart of the EMSA procedure and the associated experience-specific possibilities

commercialize systems suitable for the production of recombinant proteins in *Escherichia coli*. The protein of interest should be purified to at least 95 % as observed on a Coomassie-stained SDS-PAGE. If the purification of the protein of interest is an issue or if elaborate multipartite DNA-binding complexes are studied, crude cell lysates or nuclear extracts containing the protein(s) of interest may be a convenient choice instead. When bacterial lysate is used, the probe may need to be labeled by incorporation because active exonucleases can promote degradation of the labeled moiety at the 5′ end of the probe. Bacterial lysates can be treated with streptomycin sulfate prior to use to precipitate nucleic acids and facilitate their

elimination by centrifugation [4]. The use of lysates is a faster and cheaper option when the DNA-binding properties have to be compared between multiple biological conditions and/or treatments. Finally, when lysates are used, small probes (oligonucleotides) are better suited to minimize false positives, and antibodies should be used to confirm the identity of a given DNA-binding protein.

1.3 The Binding Reaction

The physical and chemical conditions of the binding reactions are decisive for the observation of shifted probes. Table 1 shows a non-exhaustive list of parameters that require optimization for optimal binding. The buffer, protease inhibitors, and EDTA help protect the protein of interest from proteases, while salts and stabilizing and reducing agents contribute to maintain its folding and functionality. Optimal parameters greatly vary from one DNA-binding protein to another and may be difficult to predict when information regarding the biochemical properties of the protein is limited. Preliminary assays designed to screen various parameters can maximize the odds of observing a shift. Observation of distinct species and their binding behavior in various conditions may give appreciable biochemical cues for functional inferences. Most biological interactions are relatively weak since strong affinity could fix a state. An advantage of EMSA is that protein–DNA complexes are stabilized by the cage effect or molecular sequestration [5, 6]. In other words, during migration, dissociated complexes tend to reassociate because their partners are maintained close to each other by the mesh of the gel matrix. Coherently, smears are often associated with diffusion of low-affinity complexes. Controls are crucial to validate specific binding. A mandatory negative control is the migration of the probe in the absence of the protein. Other important negative controls can be biologically relevant DNA regions not thought to be bound or addition of competitor DNA (*see* Table 1). Known targets are great positive controls, as well as supershifts using antibodies raised against the protein of interest. EMSA can provide key information about protein–DNA interactions; however, other experiments such as DNAseI footprinting, chromatin immunoprecipitation, or reporter assays are necessary to confirm binding specificity and relevance in a given biological context.

1.4 Applications

Thermodynamics. The biological activity of a transcriptional regulator or any other DNA-binding protein usually correlates with its affinity for its recognition motif(s), which is quantified as a dissociation constant (K_D) or association constant (K_A) of a protein for a DNA target in defined conditions [7]. The K_D is usually determined by titration of the labeled probe upon addition of fold increments of the protein and subsequent quantification of the unbound and protein-bound DNA by densitometry.

Table 1
Physical and chemical parameters of the binding reaction

Parameter	Options[a]
Temperature	4 °C, 25 °C, 37 °C, or higher if proteins of heat-tolerant organisms are studied
Time	5 min to 2 h
Composition of the binding buffer	• Buffer (5–50 mM) – Tris, Glycine, HEPES, MOPS, Bis-Tris, Phosphate • Salts (1–50 mM) – NaCl, KCl, $MgCl_2$, $CaCl_2$, $ZnCl_2$ • Competitor DNA (0.5–50 μg/ml) – Bulk or sonicated chromosomal DNA Salmon sperm DNA *E. coli* chromosome Calf thymus – Synthetic DNA Poly d(I-C) or Poly d(A-T) Polysulfated carbohydrate heparin Unlabeled specific probes Designed oligonucleotides • Stabilizing agents – Glycerol, sucrose, BSA, triethylene glycol, Nonidet-P40 • Protease inhibitors – PMSF, protease inhibitor cocktails, EDTA • Reducing agent – DTT (*see* **Note 3**)

[a]These parameters, which vary depending on the nature of the protein and nucleic acid binding partners, usually require optimization for optimal binding. Further information can be found in [2, 3]

Stoichiometry. The use of a mix of wild-type and truncated or lengthened mutant proteins (which exhibit wild-type binding and multimerization properties) allows an estimation of the number of monomer subunits in a protein complex [8, 9]. The mutant proteins migrate differently than their wild-type counterparts given their altered size generating additional bands in which the DNA is bound either by the mutant proteins alone or by a mix of wild-type and mutant monomers. The number of additional bands allows the approximation of the number of subunits in the DNA-binding complex. The binding of more than one protein can also be studied by EMSA [10]. Additional protein partners alter the size, conformation, and charge of the complex, generating a discernible shift in the migration pattern of the labeled probe. An EMSA-based method was even designed to determine the molecular weight of complexes [11]. Relative affinity of multiple partners can be assessed by K_D determinations in defined conditions, preferably side-by-side in the same experiment since complex formation and migration are highly influenced by the experimental conditions [12].

The number of shifted bands observed represents the equilibrium state at the moment of loading and through the gel during migration. Generally, cooperative binding to multiple sites leads to a single shifted band because the equilibrium is rapidly shifted toward the all bound state. In contrast, in the case of noncooperative binding to adjacent sites, EMSA can reveal arrays of complexes of different sizes that form between the protein of interest and its target operators.

Topology. The sequence of the DNA probe itself can disturb migration patterns. For example, repetitions of 4–6 A•T tracks in phase with the DNA helix (every 10 bp) lead to significant bending of the DNA [13, 14]. Such a deflection of the helix axis is also often observed after binding of transcription factors or DNA processing enzymes. Succession of phased bends, either attributed to the sequence or protein binding, has been used in studies investigating the topology of DNA [15]. An approximation of the bending angle can be calculated by using permutated probes, i.e., probes of identical size partially overlapping each other and containing a given operator [16, 17]. Bending can lead to DNA loops often seen for strong repressors [18] or even be crucial for the functionality of activators [19].

SELEX. The EMSA principle was brought to the next level with the development of the systematic evolution of ligands by exponential enrichment procedure (SELEX) [20, 21]. Briefly, this method screens randomized nucleic acids for their affinity for a ligand that can be a protein or a small molecule. A clear advantage of this method when screening targets of DNA-binding proteins is that the results reflect a biologically relevant recognition motif, important for the establishment of an accurate consensus sequence [22]. The automation of the method largely contributed to the field of aptamer study and discovery, and several derivatives of the methods were developed [23].

2 Materials

Prepare all solutions using analytical grade reagents and molecular grade water for enzymatic reactions or deionized water for the preparation of buffers. Protein handling should always be on ice to maintain the activity. Diligently follow all waste disposal regulations, especially for the material contaminated with toxic reagents or radioisotopes. The following material marked by an asterisk (*) is hazardous and harmful to the environment, and should be handled and discarded accordingly.

2.1 Binding Reaction Components

1. Purified protein of interest and the appropriate dilution buffer (*see* **Note 4**).
2. Reagent for protein quantification (*see* **Note 5**).

3. DNA template for amplification (genome, plasmid, or amplicon).
4. Primers to specifically amplify the probe of interest (*see* **Note 6**).
5. Taq DNA polymerase and the buffer provided by the manufacturer.
6. Commercial PCR Purification Kit or phenol–chloroform DNA extraction reagents [24].
7. T4 polynucleotide kinase and the buffer provided by the manufacturer.
8. *ATP γ-^{32}P 3000 Ci/mmol. Caution: ^{32}P emits harmful high-energy beta and secondary X radiations (*see* **Note 7**).
9. G-25 Sephadex columns.
10. Binding reaction buffer (2×): 20 mM Tris–HCl pH 8.0, 20 mM KCl, 1 mM $MgCl_2$, 5 mM DTT, 5 % glycerol. Mix 400 μl 1 M Tris–HCl pH 8.0, 1.6 ml 250 mM KCl, 10 μl 2 M $MgCl_2$, 15.42 mg DTT, and 0.5 ml glycerol. Complete the volume to 10 ml. Divide in aliquots and store at −20 °C (*see* Table 1).
11. Thermocycler.
12. Nanodrop (Thermo Scientific) or other device to quantify the DNA concentration.
13. Dry bath at 37 °C. The use of wet baths can mask a spill of radioactive material.
14. Radioactivity general equipment and permit.

2.2 Polyacrylamide Gel Components and Detection

1. 95 % or anhydrous ethanol.
2. Detergent powder for glassware.
3. Extra low-lint delicate task wipers.
4. Optional: *PlusOne Bind-Silane mixture: 3 μl γ-methacryloxypropyltrimethoxysilane (PlusOne Bind-Silane, GE Healthcare), 950 μl 95 % ethanol, 50 μl glacial acetic acid. Mix well and keep at room temperature for several months in a tightly sealed tube.
5. Tris borate EDTA (TBE; 10×): 890 mM Tris Base, 890 mM boric acid, 20 mM EDTA. 108 g Tris base, 55 g boric acid, 40 ml of 0.5 M EDTA pH 8.0. Filter to remove particles in the buffer that may affect migration. Sterilize by autoclaving and store at room temperature.
6. *Ammonium persulfate: 10 % (w/v) solution in water. Divide in single-use aliquots and store them at −20 °C.
7. *N,N,N,N-Tetramethylethylenediamine (TEMED). Store at 4 °C.
8. 40 % acrylamide/bis. Store at 4 °C.
9. Gel running buffer: 0.5× TBE. Dilute 50 ml of 10× TBE in 950 ml of deionized water. Keep at 4 °C (*see* **Note 2**).

10. Sample loading buffer: 0.25 % (w/v) bromophenol blue, 30 % (v/v) glycerol (*see* **Note 8**).
11. Fixing solution: 20 % (v/v) MeOH, 10 % (v/v) acetic acid in deionized water.
12. Filter paper.
13. Thin plastic food wrap (e.g., Saran wrap).
14. Vertical electrophoresis apparatus, ideally with a cooling system that helps stabilizing protein–DNA complexes. For example, the Emperor Penguin Water Cooled Dual Gel Electrophoresis System P9DS (Owl). Alternatively, gels can be run in a cold room.
15. Electrophoresis power supply (250 V, 200 mA capacity recommended).
16. Optional: Device for degassing the polyacrylamide gels [2].
17. Gel dryer.
18. Phosphor screen and cassette (Kodak).
19. Phosphorimager instrument (e.g., Storm 860 Molecular Imager from GMI).
20. Image analysis software (e.g., QuantityOne from Bio-Rad).

3 Methods

All procedures must be carried on ice unless otherwise indicated.

3.1 Preparation of the DNA Probe

1. Prepare 6 PCR reactions of 50 µl to amplify the probe by following the Taq polymerase manufacturer's instructions. Optimize the PCR conditions to obtain a single amplicon as observed on an ethidium bromide-stained agarose gel. Caution: Ethidium bromide is a suspected mutagenic agent.
2. Pool and purify the PCR products.
3. Measure the DNA concentration of the purified probe and convert it to pmol/µl (*see* **Note 9**).

3.2 Probe Labeling

1. Prepare the labeling reaction mixture as described in Table 2. Incubate for 1 h at 37 °C. Shorter incubation time can be used, but the efficiency of the labeling will decrease accordingly. If nucleic acid material other than amplicons is used (e.g., oligonucleotides, RNAs), ensure that it is free of 5′ phosphate; otherwise treat with the Antarctic phosphatase prior to the labeling.
2. Purify the labeled probe with a G-25 column to eliminate the unused ATP γ-^{32}P 3000 Ci/mmol. This step lowers the exposure to harmful radiations and prevents the camouflage of bands of

Table 2
DNA probe-labeling reaction mixture

Reagent	Volume (µl)
10× polynucleotide kinase buffer	1
10 pmol/µl labeled and purified probe	2
ATP γ-^{32}P 3000 Ci/mmol	1–5 (*see* **Note 7**)
T4 DNA polynucleotide kinase (10 U/µl)	1
H_2O	Up to 10

Table 3
Binding reaction mixture

Reagent	Volume (µl)
2× reaction buffer	12
Diluted protein	0.25–8
Competitor DNA[a]	0.25–8
Diluted labeled probe	1
H_2O	Up to 24 µl

[a]The concentration and the nature of the competitor DNA depend on the nucleic acid sample and the affinity of the protein of interest for the targets

low molecular weight by the strong signal generated by unused radioactive ATP.

3. Dilute the probe to an appropriate concentration in molecular grade water (*see* **Note 10**).

3.3 Binding Reaction

1. Just before use, dilute the protein sample with the appropriate buffer to concentrations ranging from 1 nM to 4 µM. Beware: Diluted proteins tend to rapidly loose activity over time.
2. Set the binding reactions on ice (without the labeled probe) and incubate for 10 min at room temperature (*see* Table 3). This step allows binding of the protein to competitor DNA. Remember to always prepare a sample without protein as a negative control.
3. Add the labeled probe. Incubate for 30 min at 37 °C, then 5 min on ice to stabilize the complexes.
4. Add the loading buffer to the sample and keep on ice until gel loading.

3.4 Separation of Protein–DNA Complexes with a Native Polyacrylamide Gel

1. Carefully wash the glass plates, spacers, and combs with detergent and hot water. Rinse with deionized water, followed by 95 % ethanol. Dry the glass plates with delicate task wipes to retrieve any remaining dust and particles. Optional: Treat one glass plate with 50 µl of PlusOne Bind-Silane mixture only in the area that will contact the wells to prevent breaking of the gel when retrieving the comb.
2. Assemble the glass plates–spacers sandwich (*see* **Note 11**).
3. Prepare the acrylamide gel as depicted in Table 4 (*see* **Note 12**). Mix thoroughly the H_2O, 2× TBE, and acrylamide by swirling gently to avoid introducing bubbles. An optional degassing step of the gel mixture facilitates acrylamide polymerization and provides a more homogenous gel. Ensure to keep a small volume of gel mixture (≈1 ml) as a first polymerization control. Caution: Non-polymerized acrylamide is a potent neurotoxic agent.
4. Just before pouring, add 500 µl of 10 % APS and 28 µl of TEMED and mix thoroughly by swirling gently to avoid introducing bubbles. Cold reagents or lower amounts of APS and TEMED will slow down the polymerization reaction, which might be useful for beginners.
5. Pour the acrylamide gel and insert the comb. Make sure to avoid bubbles in the gel which will lead to distorted migration patterns (*see* **Note 13**).
6. Once the acrylamide is polymerized, assemble the electrophoretic apparatus with the gel(s) and the migration buffer (0.5× TBE).
7. Precool the gel at 4 °C for 1 h prior to migration. The gel can be stored at 4 °C up to 1 week in 100 % humidity in a sealed bag with a wet paper.

Table 4
Preparation of native polyacrylamide gels of various concentrations

Reagent	4 %	6 %	8 %	10 %	15 %
H_2O (ml)	22.75	21	19.25	17.5	13.12
2× TBE[a] (ml)	8.75	8.75	8.75	8.75	8.75
40 % acrylamide (ml)	3.5	5.25	7	8.75	13.13
10 % APS[b] (µl)	500	500	500	500	500
TEMED[b] (µl)	28	28	28	28	28
H_2O (ml)	22.75	21	19.25	17.5	13.12

[a]Volume for a final concentration of 0.5× TBE. Final concentration can range from 0.25× to 1×. Higher concentrations increase the speed of migration but also generate heat (*see* **Note 2**)
[b]APS and TEMED should be added just prior to pouring as they trigger polymerization of the acrylamide

8. Just before use, remove the comb and wash carefully the wells to remove non-polymerized acrylamide. It might be helpful to draw the edges of the wells on the glass plate if they are difficult to see after the removal of the comb to facilitate loading of the samples.

3.5 Migration and Detection

1. Pre-run the gel at 135 V (10 V/cm gel length). A second polymerization control can be carried out at this step by loading 2 μl of loading buffer (without sample) in every well. The migration of the bromophenol blue indicates if polymerization occurred properly. If not, discard the gel and prepare a new one.
2. Stop the power supply and quickly load the samples. After loading, run the gel at 120 V (*see* **Note 14**).
3. After migration, disassemble the apparatus and dry the gel plates assembly. Residual buffer can lead to vacuum formation when separating the plates, and contact of free buffer on the gel can elute radioactive nucleic acids potentially increasing the background signal in the autoradiography. Separate the glass plates, leaving the gel on one plate. Put the gel (and glass plate) in a large Pyrex dish containing the fixing solution for 20–30 min.
4. Remove the glass plate from the fixing solution and transfer the gel onto a filter paper. Put the paper on the gel and flip upside down. Delicately detach the paper bound gel from the glass plate. Place the gel paper on a larger filter paper onto the gel dryer surface, the gel facing up. Cover the gel with thin plastic food wrap while making sure to avoid bubbles and wrinkles. Beware: Transfer of gels with high acrylamide content might be challenging as they tend to not stick to the filter paper. A razor blade may be useful to detach the gel.
5. Dry the gel for 90 min with heat (70 °C).
6. Expose the thin plastic food wrap-covered gel on a phosphor screen in a cassette. Bands will appear after 30 min to several days of exposition depending on the signal strength. Beware of residual moisture as it can damage the phosphor screen. Also, phosphor screens should be handled under a dim light.
7. Detect the signal using a phosphorimager.
8. Quantitative measure of signal strength can be obtained by densitometry using image analysis software.

4 Notes

1. For concerns related to health, environment, or lack of experience, nonradioactive probe labeling can be achieved using high-sensitivity DNA detection dyes such as SYBR.

The downsides of this approach in comparison to radiolabeling are its lower sensitivity and the requirement for higher quantity of biologically active material. If this is not an issue, an elegant alternative to radiolabeling is the differential staining of both DNA and protein molecules using commercially available system such as the EMSA Kit with SYBR® Green & SYPRO® Ruby EMSA stains from Life Technologies. Keep in mind that the detection of these fluorescent dyes is only possible with adapted camera filters.

2. Low ionic strength buffers (e.g., TBE, TAE, TE) contribute to the detection of complexes by increasing the speed of migration and by generating less heat. Higher ionic strength conditions can disrupt ionic bonds which are important in complex stabilization [3]. Cofactors might be needed in the binding and running buffers for specific binding. For example, addition of cAMP is necessary for optimal DNA binding by CAP [25].
3. Dithiothreitol (DTT) is used at a concentration of 1 % (w/v) to reduce disulfide bonds between subunits of proteins during electrophoresis. At lower concentrations (1 mM), DTT is used to counteract the oxidation of proteins and preserve biological activity.
4. The dilution buffer depends on the nature of the starting protein material. It is generally the buffer used during the purification steps. To retain the DNA-binding activity, protein samples should always be handled on ice and used fresh, especially for crude cell lysates. Since the freezing/thawing procedure is inherently deleterious for protein activity, purified protein samples can be kept at 4 °C in a buffer containing protease inhibitors and a reducing agent (e.g., DTT) to prevent oxidation. For cell extracts and long-term conservation of purified proteins (>2 weeks), samples should be separated in single-use aliquots, quickly frozen in liquid nitrogen and stored at −80 °C. The addition of glycerol (50 % v/v final) may help retain the activity. However, dialysis will be mandatory to remove the glycerol prior to binding assays.
5. The easiest way to determine the protein concentration of a sample is with the Layne formula:

$$\text{Protein concentration (mg/ml)} = (1.55 \times \text{OD}_{280}) - (0.76 \times \text{OD}_{260}).$$

For K_D determination or other quantitative assessments, we strongly advise the use of a more accurate method like the Bradford or BCA protein assays, respectively, commercialized by Bio-Rad and Thermo Scientific Pierce. The choice of the method depends on the composition of the protein sample as there are chemicals known to interfere with the reagents. Moreover, the composition in amino acids of the calibration protein should be considered to maximize the accuracy of the concentration measurement.

6. Ideally, the probe size should be between 50 and 600 bp. Labeled primers can also be used. Hybridized oligonucleotides are convenient to screen large libraries of mutants or single operators that are closely located in their native locus.
7. The volume of ATP γ-^{32}P 3000 Ci/mmol added to the labeling reaction can vary depending on the availability of the isotope and the specific activity of the probe needed (e.g., single complex formed, multiple complex formed). A higher concentration of radiolabeled ATP favors the labeling reaction by the T4 polynucleotide kinase. Always order and use a minimum of radioactive material because of its short-lived activity, its hazardousness, and its environmental implications. The lower tank of the electrophoresis apparatus may get contaminated with radioactive material. You should always verify with a Geiger counter and discard it accordingly.
8. In some instances, the bromophenol blue dye can disturb the binding reaction. To bypass this problem and still track the migration of the samples, the dye can be added to a well next to the undyed samples.
9. Example of conversion of double-stranded DNA concentration from ng/µl to pmol/µl:

   Probe of 100 bp after purification: 30 µl at 200 ng/µl

   $$\text{pmol}/\mu\text{l} = (0.2\ \mu\text{g}/\mu\text{l}) \times (1\ \text{pmol}/660\ \text{pg}) \times (10^6\ \text{pg}/1\ \mu\text{g}) \times (1/100\ \text{bp}) = 3\ \text{pmol}/\mu\text{l}.$$

10. The concentration of the DNA probe used can vary according to the requirements of the experiment. In the conditions indicated in the "Methods" section, a 1/50 dilution is used for <1 week after the calibration date of ^{32}P activity given by the manufacturer and 1/20 for 1–2 weeks old ^{32}P. It is better not to use material older than 2 weeks due to the radiolytic degradation of the sample.
11. The left, right, and bottom edges of the glass plates-separators sandwich can be secured with masking tape to prevent leakage of non-polymerized acrylamide gel.
12. A concentration of 8 % polyacrylamide is a good starting point when studying the binding of a small transcriptional regulator (>100 kDa) on a DNA probe of 80–300 bp. If larger proteins or DNAs are studied, the concentration can be lowered down to 4 %. A lower polyacrylamide content increases the speed of migration. Higher concentration (up to 15 %) can help separate multiple species and stabilizes complexes at a certain extent.
13. Beware when putting your face close to the casting apparatus when inserting the comb. Splashes of non-polymerized acrylamide can occur. We find convenient to pour the gel with a

60 ml syringe topped with a 20G1 needle (BD PrecisionGlide). The volume of the syringe and the gauge of the needle vary according to the volume of gel needed and the space between the glass plates. The presence of bubbles in the gel affects the migration because air does not conduct electric current well. Tilting the casting assembly at a 45° angle may help prevent bubbles formation. Gently tapping the assembly may help dislodge air bubbles.

14. It may be useful to leave an empty well between two samples to show the position of your samples on the autoradiography. Avoid expelling bubbles while loading the samples into the wells.

Acknowledgement

This work was supported by the Fonds Québécois de la recherche sur la nature et les technologies (D.P.L.) and a Discovery Grant and Discovery Acceleration Supplement from the Natural Sciences and Engineering Council of Canada (V.B.). V.B. holds a Canada Research Chair in molecular bacterial genetics.

References

1. Garner MM, Revzin A (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic Acids Res 9:3047–3060
2. Hellman LM, Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein–nucleic acid interactions. Nat Protoc. doi:10.1038/nprot.2007.249
3. Lane D, Prentki P, Chandler M (1992) Use of gel retardation to analyze protein-nucleic acid interactions. Microbiol Rev 56:509–528
4. Oxenburgh MS, Snoswell AM (1965) Use of Streptomycin in the Separation of Nucleic Acids from Protein in a Bacterial Extract. Nature. doi:10.1038/2071416a0
5. Cann JR (1989) Phenomenological theory of gel electrophoresis of protein-nucleic acid complexes. J Biol Chem 264:17032–17040
6. Vossen KM, Fried MG (1997) Sequestration stabilizes lac repressor-DNA complexes during gel electrophoresis. Anal Biochem 245:85–92
7. Fried M, Crothers DM (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res 9:6505–6525
8. Hope IA, Struhl K (1987) GCN4, a eukaryotic transcriptional activator protein, binds as a dimer to target DNA. EMBO J 6:2781–2784
9. Kimsey HH, Waldor MK (2003) The CTX Repressor RstR Binds DNA Cooperatively to Form Tetrameric Repressor-Operator Complexes. J Biol Chem 279:2640–2647
10. Fried MG, Daugherty MA (1998) Electrophoretic analysis of multiple protein-DNA interactions. Electrophoresis 19:1247–1253
11. Orchard K, May GE (1993) An EMSA-based method for determining the molecular weight of a protein–DNA complex. Nucleic Acids Res 21:3335
12. Carey MF, Peterson CL, Smale ST (2013) Electrophoretic mobility-shift assays. Cold Spring Harb Protoc. doi:10.1101/pdb.prot075861
13. Haran TE, Mohanty U (2009) The unique structure of A-tracts and intrinsic DNA bending. Q Rev Biophys. doi:10.1017/S0033583509004752
14. Koo HS, Wu HM, Crothers DM (1986) DNA bending at adenine. thymine tracts. Nature 320:501–506

15. Zinkel SS, Crothers DM (1987) DNA bend direction by phase sensitive detection. Nature 328:178–181
16. Kim J, Zwieb C, Wu C, Adhya S (1989) Bending of DNA by gene-regulatory proteins: construction and use of a DNA bending vector. Gene 85:15–23
17. Wu HM, Crothers DM (1984) The locus of sequence-directed and protein-induced DNA bending. Nature 308:509–513
18. Griffith J, Hochschild A, Ptashne M (1986) DNA loops induced by cooperative binding of lambda repressor. Nature 322:750–752
19. Liu-Johnson HN, Gartenberg MR, Crothers DM (1986) The DNA binding domain and bending angle of *E. coli* CAP protein. Cell 47:995–1005
20. Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346:818–822
21. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505–510
22. Lee Y-Y, Barker CS, Matsumura P et al (2011) Refining the Binding of the *Escherichia coli* Flagellar Master Regulator, $FlhD_4C_2$, on a Base-Specific Level. J Bacteriol 193: 4057–4068
23. Stoltenburg R, Reinemann C, Strehlitz B (2007) SELEX–a (r)evolutionary method to generate high-affinity nucleic acid ligands. Biomol Eng 24:381–403
24. Sambrook J, Russel DW (2001) Molecular cloning : A laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, New York
25. Fried MG, Crothers DM (1984) Equilibrium studies of the cyclic AMP receptor protein-DNA interaction. J Mol Biol 172:241–262

Chapter 2

In Vitro DNase I Footprinting

Benoît P. Leblanc and Tom Moss

Abstract

The association of proteins with the DNA double helix can interfere with the accessibility of the latter to nucleases. This is particularly true when using bulky nucleases such as DNase I. The DNase I footprinting method was developed to take advantage of this fact in the study of DNA-protein interactions: it consists in comparing the pattern of fragments generated by the partial digestion of a DNA sequence in the absence of a protein to that produced by its partial digestion in the presence of said protein. Normally, when the two sets of fragments are separated side by side on a gel, the ladder of DNase I-generated fragments produced in the presence of the protein will feature blank regions (devoid of fragments, indicating protection) and/or enhanced cleavage sites (indicating increased availability to the nuclease). This technique can furthermore reveal if multiple sites for a DNA-binding protein are present on a same fragment and in such a case will also allow the comparison of their respective affinities.

Key words DNase I, Footprinting, Protection, Sequencing

1 Introduction

DNase I footprinting was developed by David Galas and Albert Schmitz in 1978 as a method to study the sequence-specific binding of proteins to DNA [1]. In the technique, a suitable uniquely end-labeled DNA fragment is allowed to interact with a given DNA-binding protein. The protein-DNA complex is then partially digested with DNase I. The protein bound to DNA protects the sequence with which it interacts from the attack by the nuclease, and subsequent molecular weight analysis of the degraded DNA by electrophoresis and autoradiography identifies the region of protection as a gap in the otherwise continuous background of digestion products; for example, *see* Fig. 1. The technique can be used to determine the site of interaction of most sequence-specific DNA-binding proteins but has been most extensively applied to the study of transcription factors. Because the DNase I molecule is relatively large as compared to other footprinting agents, its attack on the DNA is sterically hindered by the presence of large molecules

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_2,

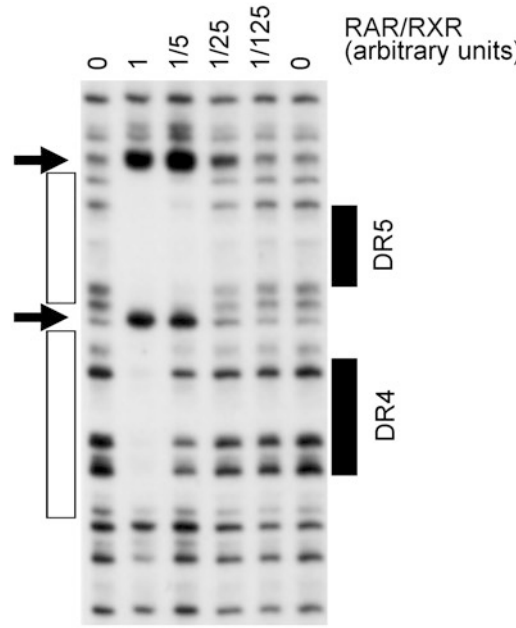


Fig. 1 Example of DNase I footprinting. A DNA fragment containing a mutated version of the mouse RARβ2 gene promoter, labeled downstream of the initiation site, was incubated with decreasing amounts of recombinant RAR/RXR heterodimers and partially digested with DNase I. The direct repeat (DR) elements bound by the proteins are indicated by *solid boxes*. Protected regions are indicated by *empty boxes*. DNase I-hypersensitive sites, which are more frequently cleaved in the presence of the proteins than in their absence, are indicated by *arrowheads*. Note that the heterodimers bind to direct repeats separated by 5 nucleotides (DR5) with roughly five times more affinity than to direct repeats separated by 4 nucleotides (DR4), a preference that can be easily visualized by this technique

such as polypeptides. Thus, DNase I footprinting is the most likely of all the footprinting techniques to detect a specific DNA-protein interaction. This is clearly demonstrated by our studies on the transcription factor xUBF, a factor binding DNA with relatively low affinity (*see* Fig. 2).

DNase I footprinting can not only be used to study the DNA interactions of purified proteins but also as an assay to identify proteins of interest within a crude nuclear extract (e.g., *see* ref. 2) Thus, it can serve much the same function as an EMSA analysis in following a specific DNA-binding activity through a series of purification steps. DNase I footprinting can often be used for proteins that do not perform well in EMSA experiments (the abovementioned UBF is an example of a protein more easily studied by footprinting than by EMSA) and so can be seen as a viable alternative. Its resolution is also greater than that of an EMSA in that it can distinguish between multiple, non-contiguous binding sites on the

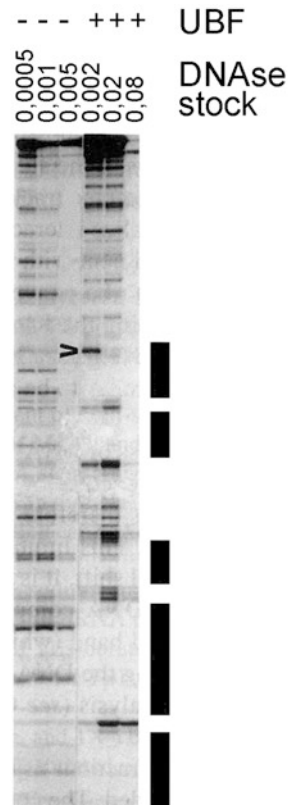


Fig. 2 Course of digestion with increasing amounts of DNase I. Here, xUBF was footprinted on the *Xenopus* ribosomal promoter using a 5′ end-labeled fragment. The numbers above the tracks refer to the DNase I dilution employed (in units/μL), and minus and plus signs refer to the naked and complexed DNAs, respectively. The predominant footprints are indicated by *solid boxes*. A visible double-stranded fragment, which could be mistaken for a hypersensitive site, is marked by an *arrowhead*

same DNA fragment. However, because of the need of an excess of protein to generate a clear footprint, the technique requires considerably more material than would an EMSA assay and cannot distinguish individual components of heterogeneous DNA-protein complexes.

DNase I (E.C.3.1.4.5) is a protein approximately 40 Å in diameter. It binds in the minor groove of the DNA helix and cuts the phosphodiester backbone of both strands independently, leaving a nick [3]. Its bulk helps to prevent it from cutting the DNA under and around a bound protein. However, a bound protein will also usually have other effects on the normal cleavage by DNase I (*see* Figs. 1 and 2). It is also not so uncommon to observe a change in the pattern of DNase cleavage without any obvious extended protection (e.g., *see* Fig. 2).

Unfortunately, DNase I does not cleave the DNA indiscriminately, some sequences being very rapidly attacked while others remain unscathed even after extensive digestion [4]. This results in a rather uneven "ladder" of digestion products after electrophoresis, something which limits the resolution of the technique (as seen in the naked DNA tracks in Figs. 1 and 2). However, when the protein-protected and naked DNA ladders are run alongside each other, the footprints are normally quite apparent. To localize the position of the footprints, sequencing ladders originating from the same nucleotide that is labeled in the footprinted DNA fragment should accompany the naked and protected tracks (*see* **Note 1**). As a single end-labeled fragment allows one to visualize interactions on only one strand of DNA at a time, it is informative to repeat the experiment with the same fragment labeled on the other strand. DNA fragment can be conveniently 5′-labeled using T_4 DNA polynucleotide kinase and 3′-labeled using the Klenow or the T_4 DNA polymerases (in fill out reactions) or terminal transferase (e.g., *see* ref. 6). To analyze both strands of the DNA duplex side by side and make direct comparisons between the two, two footprinting reactions should be run in parallel. Both reactions would use an equivalent DNA fragment, labeled at the same end, but in one case, the 5′ strand would be labeled while in the other it would be the 3′ strand.

DNase I footprinting requires an excess of DNA-binding activity over the amount of DNA fragment used. The higher the percent occupancy of a site on the DNA, the clearer the observed footprint will be. It is therefore important not to titrate the available proteins with too much DNA. This limitation can, in part, be overcome when a protein also generates a gel shift. It is then feasible to fractionate the partially DNase-digested protein-DNA complex by nondenaturing gel electrophoresis and to excise the shifted band (which is then a homogeneous protein-DNA complex) before analyzing the DNA by denaturing gel electrophoresis as in the standard footprint analysis.

Footprinting crude or impure protein fractions usually require that an excess of a nonspecific competitor DNA be added to the reaction. The competitor binds nonspecific DNA-binding proteins as effectively as the specific labeled target DNA fragment and hence, when present in sufficient excess, leaves the main part of the labeled DNA available for the sequence-specific protein. Homogeneous and highly enriched protein fractions usually do not require the presence of a nonspecific competitor during footprinting. When planning a footprinting experiment, it is prerequisite to start by determining the optimal concentration of DNase I to be used. This will be a linear function of the amount of nonspecific competitor, but more importantly (although less reproducibly), it will also be a function of the amount and purity of the

protein fraction added. As a general rule, more DNase I will be required if more protein is present in the binding reaction, whether or not this protein binds specifically. Thus, very different DNase concentrations may be required to produce the required degree of digestion on naked and protein-bound DNA. A careful titration of the DNase concentration is therefore essential to optimize the detection of a footprint and can even make the difference between the detection of a given interaction or lack thereof.

The following protocol was developed to study the footprinting of the *Xenopus* ribosomal transcription factor xUBF, which is a rather weak DNA-binding protein with a broad specificity. The protocol has been used successfully for studies on the human retinoic acid receptor alpha and the yeast Ace1 transcription factor, among others. We recommend that the reader also refer to ref. 5 for more information on the quantitative analysis of protein-DNA interactions by footprinting.

2 Materials

1. Binding buffer (2×): 20 % glycerol, 0.2 mM EDTA, 1 mM dithiothreitol (DTT), 20 mM HEPES pH 7.9, and 4 % poly(vinyl alcohol) (*see* **Note 2**).
2. Poly(dA-dT)-poly(dA-dT) (Sigma-Aldrich): 1 mg/mL solution in TE. Keep at −20 °C (*see* **Note 3**).
3. End-labeled DNA fragment of high specific activity (*see* **Note 1**).
4. Cofactor solution: 10 mM $MgCl_2$, 5 mM $CaCl_2$.
5. DNase I stock solution: A standardized vial of DNase I (Sigma-Aldrich #D4263) is dissolved in 50 % glycerol, 135 mM NaCl, 15 mM CH_3COONa pH 6.5 at a concentration of 10 U/μL. This stock solution can be kept at −20 °C indefinitely (*see* **Note 4**).
6. 1 M KCl solution.
7. Reaction stop solution: 1 % sodium dodecyl sulfate (SDS), 200 mM NaCl, 20 mM EDTA pH 8.0, and 40 μg/mL tRNA (*see* **Note 5**).
8. 10× TBE buffer: 900 mM Tris-borate pH 8.3 (108 g/L Tris base and 55 g/L boric acid, 20 mM EDTA).
9. Loading buffer: 95 % formamide, 0.05 % xylene cyanol, 0.05 % bromophenol blue.
10. 6 % acrylamide, 7 M urea, and 1× TBE sequencing gel.
11. Phenol-chloroform-isoamyl alcohol (25:24:1, v/v).
12. Sequenase 2.0 DNA sequencing kit (Affymetrix #70770) or an equivalent system.

13. A Nunc MicroWell MiniTray, also called Terasaki plate (Sigma-Aldrich #M0815) (*see* **Note 9**).
14. A sequencing primer of appropriate sequence, as defined in **step 12** of the "Methods" section.
15. Fixing solution: 10 % ethanol, 10 % acetic acid.
16. Blotting paper (Whatman 3MM or the equivalent).

3 Methods

The footprinting reaction is performed in three stages: (1) binding of the protein to the DNA, (2) partial digestion of the protein-DNA complex with DNase I and recovery of these fragments, and (3) separation of the fragments on a DNA sequencing gel.

1. Each binding reaction is performed in a total volume of 50 µL containing 25 µL of 2× binding buffer, 0.5 µL of 1 mg/mL poly(dA-dT)-poly(dA-dT), 2–3 ng of end-labeled DNA fragment (approximately 15,000 CPM) (*see* **Note 6**), an adjustable volume of protein fraction, and enough 1 M KCl to bring the final KCl concentration to 60 mM. The maximum volume of the protein fraction that can be used will often be determined by how much salt it contains. Each reaction is performed in a 1.5 mL Eppendorf tube. The following work chart can be used to facilitate the experiment.

Tube #	Labeled DNA (15,000 CPM)	2× Binding buffer	Poly(dA-dT)-poly(dA-dT) 1 mg/mL	Protein fraction	1 M KCl (to 60 mM final)	H_2O	Final volume
	µL	25 µL	0.5 µL	µL	µL	µL	50 µL
1		25 µL					50 µL
2		25 µL					50 µL
3		25 µL					50 µL
4		25 µL					50 µL
…		25 µL					50 µL

2. Incubate on ice for 20 min (binding reaction).
3. During the binding reaction, prepare the many DNase I working dilutions in 1.5 mL Eppendorf tubes. We suggest diluting an aliquot of the stock solution in water, on ice, to concentrations ranging from 0.0005 to 0.1 Kunitz units/µL. A good range of useful dilutions would consist in aliquots at 0.0005, 0.001, 0.002, 0.004, and 0.008 for the digestion of naked

DNA and 0.005, 0.01, 0.02, 0.04, and 0.08 for the digestion of protein-DNA complexes (*see* **Note 7**).

4. Set up the digestion reaction. You should have three micropipettes on hand: the first one set at 50 µL, another set at 5 µL, and the last one set at 100 µL. Pipette tip boxes should be nearby and kept open (as **step 5** will require to move quickly). The cofactor solution and the reaction stop solutions should both be at room temperature (R/T), in open tubes, and within easy reach. The DNase I dilutions should also be within easy reach and kept on ice, with the tube lids open. Have a watch or a stopwatch on the bench (not on your wrist). Since the DNA fragment is radioactive, all the operations should be performed behind a plexiglass shield.
5. Once the 20 min of the binding reaction is over, transfer the reaction tubes, eight at a time, to a rack at R/T. Each of the eight tubes will be processed one after the other, at 15 s intervals. Start with the first tube and (A) add 50 µL of cofactor solution; (B) add 5 µL of the appropriate DNase I dilution; (C) cap the tube and move on to the next one. (Roughly 15 s should have elapsed, as you can judge from the watch on the bench.) Proceed to the second tube, and repeat again for the following ones.
6. After having processed the eighth tube, go back to the first one and stop the digestion reaction by adding 100 µL reaction stop solution. As previously, move on to each subsequent tube every 15 s and add 100 µL reaction stop solution to them. The total digestion time for each of the eight tubes will have been two minutes (*see* **Note 8**).
7. After all the reactions have been processed, extract each of them with one volume of the phenol-chloroform mix and transfer the aqueous phase of each tube to a fresh one.
8. Add at least one volume of isopropanol to the extracted aqueous phase to precipitate the nucleic acids.
9. Microcentrifuge for 15 min at approximately 15,000×*g*. Remove the supernatant with a micropipette and keep it in a fresh tube in case the nucleic acids did not precipitate. Check the presence of a radioactive pellet with a Geiger counter. (The pellet might come unstuck and could be found floating in the supernatant. If that occurs, centrifuge again.)
10. Add 100 µL ice cold, 70 % ethanol. Microcentrifuge for two minutes as above. Remove the supernatant, checking with the Geiger counter that you did not pick up the pellet (the supernatant at this step should not be radioactive, or only very slightly). Air-dry the pellets or briefly dry in a vacuum desiccator. Do not overdry.

11. Resuspend each pellet in 5 μL loading buffer, vortex, and centrifuge briefly. The samples are now ready to be loaded on a sequencing gel but can also be kept at −20 °C and run at a later time.
12. A sequencing ladder should be run in parallel with the samples on the sequencing gel (**step 13**). In the early years of DNase I footprinting, this ladder was mostly a G+A chemical degradation ladder generated by the Maxam and Gilbert method using the same end-labeled fragment as for the footprinting reaction itself [6]. However, a better resolution can be achieved by using the four tracks (A, C, G, and T) of a sequencing reaction performed according to the Sanger protocol [7]. This can be done using homemade reagents or a commercial kit such as the one listed in the "Materials" section. For the sequencing ladder and the footprinting ladders to be properly aligned, care must be taken in using a sequencing primer that aligns precisely with the labeled base of the DNA fragment that is footprinted (*see* **Note 9**). Although such a sequencing reaction is performed with ^{35}S instead of ^{32}P, the intensity of its signal after autoradiography is quite acceptable and in the range of that of the footprint ladders.
13. Pre-run a standard 6 % polyacrylamide sequencing gel until hot (50–55 °C, roughly 30 min) before loading the samples. Wash the wells thoroughly before loading. Denature the samples for two minutes at 95 °C and load with a micropipette. (We find that regular tips, wedged at the top of the wells, work just as well as specialty elongated tips and do not clog as easily). Run the gel hot to keep the DNA denatured (*see* **Note 10**). The gel should be run until the xylene cyanol reaches 2/3 of its way to the bottom.
14. After the run, fix the gel with 10 % ethanol and 10 % acetic acid. It is not necessary to immerse the gel in the solution; laying it flat in or near a sink and covering it with a thin layer of solution will suffice. (The gel plate can be put in a large tray if no large sink is available.) The solution can be replenished from time to time for about 5–10 min. The gel should then be covered with a wet piece of blotting paper and washed with gently running water for about 3–4 min, which will help remove the urea. The paper is gently removed (make sure the gel doesn't stay attached to it or start tearing up!) and replaced with a new, dry piece of blotting paper. The gel is then dried in a gel drier and wrapped in Saran.
15. Expose the gel overnight to an autoradiography film or to a phosphorimager plate.

4 Notes

1. Single-stranded breaks in the end-labeled DNA fragment must be avoided as they give false signals indistinguishable from genuine DNase I cleavage and hence can mask an otherwise good footprint. It is therefore advisable, when running the gel, to reserve one lane for the undigested footprinting probe. This should give only one band (uncleaved and undegraded). Radiochemical nicking is a theoretical threat to the probe's integrity, but in practice, it has never caused us any problem.
2. This binding buffer has been shown to work well for the transcription factor NF-1 [5]. In our hands, it has also worked well for several other factors, such as the human retinoid receptors RAR and RXR, the yeast Ace1 factor, and *Xenopus* and human UBF. Glycerol and poly(vinyl alcohol) (an agent used to reduce the available volume of water and hence concentrate the binding activity) are not mandatory. In fact, very simple binding buffers can be used; the original footprinting conditions for the binding of the *lac* repressor to the *lac* operator were 10 mM cacodylate buffer pH 8.0, 10 mM $MgCl_2$, 5 mM $CaCl_2$, and 0.1 mM DTT [1]. Particular conditions of binding might be required depending on the protein studied.
3. Poly dIdC is another popular nonspecific general competitor. Its only drawback is that it may act as a specific competitor for proteins favoring GC-rich stretches. Another nonspecific competitor can be sheared genomic DNA.
4. Note that the usual unit for DNase I activity is the *Kunitz* unit; the name "Kunitz" should not be mistaken for one thousand units. The standardized vials give very reproducible results. Glycerol in the buffer will keep the enzyme from freezing.
5. Do not be tempted to use too much carrier tRNA, as an excess of it causes a very annoying fuzziness of the gel bands, preventing resolution of closely packed individual bands. If attempting to omit the carrier altogether, make sure that the labeled DNA does precipitate by using the Geiger counter. Alternatively, 10 μg glycogen can be used as a carrier.
6. The use of 5′ end labeling with polynucleotide kinase in the presence of crude protein extracts can sometimes lead to a severe loss of signal because of the presence of phosphatases. In these cases, 3′ end labeling by "filling out" with Klenow or T_4 DNA polymerase is to be preferred.
7. Ranges of dilutions should be tested empirically with the experimenter's material. Initial experiments should be only concerned with defining the appropriate range of dilutions that give the best looking footprint. In **step 3**, we suggested a series of twofold dilutions; a series of fivefold digestions would

also be acceptable (and require fewer lanes on a gel, as well as less material) to identify the best-suited range. Note that naked DNA or DNA-protein complexes using highly purified proteins will always require less DNase than when crude protein fractions are used.

8. The 15 s intervals are quite convenient once one gets the hang of the technique. Intervals of 10 s can be considered for the most daring, but they increase the chance of making a mistake and do not really save time. For beginners, intervals of 30 s should be considered (and so four tubes could be processed at a time instead of eight).
9. The chemical G + A ladder has the advantage of using the same labeled DNA fragment as the footprinting reaction; however, it requires the use of toxic reagents, and its interpretation can be ambiguous since it does not differentiate between purines. The four ladders generated by the dideoxynucleotide method allow a full read of the region of interest, use mostly innocuous reagents, and have a long shelf life since they are labeled with ^{35}S. For them to be aligned with the footprint ladders, the only thing to pay attention to is that the 5′ end of the sequencing primer be at the same position as the labeled nucleotide of the footprinting probe. For example, let us imagine a DNA fragment cut with *Eco*R I and labeled with T_4 kinase. The labeled fragment (the one that is visible on the autoradiogram of the gel) would look like this: 5′ *AATTCNNNNNN… 3′, where the asterisk represents the radioactive label. To match such a footprinting probe, the sequencing reaction should be performed with a primer starting at the same base: 5′ AATTCNNNNNN 3′. The elongation part of the sequencing reaction can be performed in a Terasaki plate instead of four independent tubes, using one well for each dideoxynucleotide elongation mix (*see* Fig. 3). By avoiding the constant opening and recapping of Eppendorf tubes, this makes adding the reagents much faster and efficient. After the sequencing reaction has been halted by the addition of stop/loading buffer to each well, the plate can be sealed with Parafilm and kept for many weeks. Denaturation of the DNA prior to loading on the gel is performed either by floating the plate for two minutes on a 95 °C water bath (a risky maneuver) or by laying it on a wet piece of tissue set on a dry heating block at 95 °C.
10. Sequencing gels are not denaturing unless run hot (7 M urea produces only a small reduction in the *Tm* of DNA). A double-stranded form of the full-length DNA fragment is therefore often seen on the upper part of the autoradiogram (*see* Fig. 2), especially at low levels of DNase I digestion, and can often be misinterpreted as a hypersensitive cleavage site. Comparison to an undigested track of DNA should clarify the point.

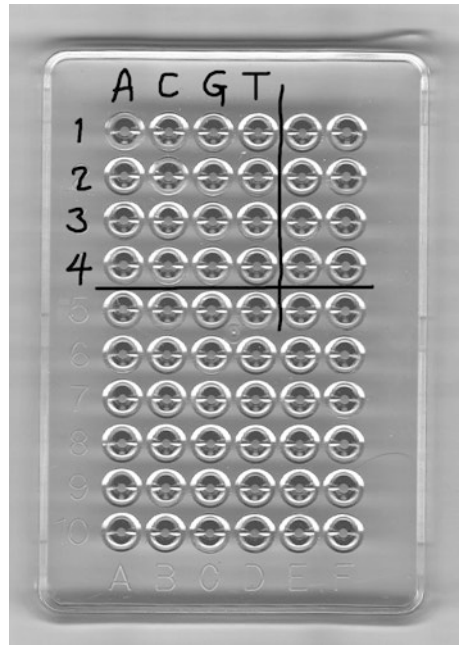


Fig. 3 A Terasaki plate used for the elongation/termination part of the Sanger sequencing reaction. An appropriate volume of each of the dideoxynucleotide elongation mixes is pipetted in the corresponding well (A, C, G, or T). To start the elongation reaction, the appropriate volume of each labeling reaction is added to each well, and incubation proceeds according to the manufacturer's instructions (usually 5 min at 37 °C). In this figure, four samples (labeled 1–4) are sequenced in parallel. The reaction is stopped according to the manufacturer's instructions, usually by addition of 5 μL formamide-bromophenol blue-xylene cyanol loading buffer. The plate makes the reaction much easier to perform, as there are no tube lids to constantly pop open and to close again

References

1. Galas DJ, Schmitz A (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. Nucleic Acids Res 5: 3157–3170
2. Rousseau S, Renaud J, Ruiz-Carrillo A (1989) Basal expression of the histone H5 gene is controlled by positive and negative cis-acting sequences. Nucleic Acids Res 17:7495–7511
3. Suck D, Lahm A, Oefner C (1988) Structure refined to 2A of a nicked DNA octanucleotide complex with DNase I. Nature 332:464–468
4. Drew HR (1984) Structural specificities of five commonly used DNA nucleases. J Mol Biol 176: 535–557
5. Brown T (1987) Analysis of RNA by Northern and slot blot hybridisation. In: Ausubel F, Brent R, Kingston R, Moore D, Seidman J, Smith J, Struhl K (eds) *Current protocols in molecular biology*. Green Publishing Associates, New York
6. Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65:499–560
7. Sanger F, Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol 94: 441–448

Chapter 3

Determining the Architecture of a Protein–DNA Complex by Combining FeBABE Cleavage Analyses, 3-D Printed Structures, and the ICM Molsoft Program

Tamara James, Meng-Lun Hsieh, Leslie Knipling, and Deborah Hinton

Abstract

Determining the structure of a protein–DNA complex can be difficult, particularly if the protein does not bind tightly to the DNA, if there are no homologous proteins from which the DNA binding can be inferred, and/or if only portions of the protein can be crystallized. If the protein comprises just a part of a large multi-subunit complex, other complications can arise such as the complex being too large for NMR studies, or it is not possible to obtain the amounts of protein and nucleic acids needed for crystallographic analyses. Here, we describe a technique we used to map the position of an activator protein relative to the DNA within a large transcription complex. We determined the position of the activator on the DNA from data generated using activator proteins that had been conjugated at specific residues with the chemical cleaving reagent, iron bromoacetamidobenzyl-EDTA (FeBABE). These analyses were combined with 3-D models of the available structures of portions of the activator protein and B-form DNA to obtain a 3-D picture of the protein relative to the DNA. Finally, the Molsoft program was used to refine the position, revealing the architecture of the protein–DNA within the transcription complex.

Key words 3-D printing, Molsoft, FeBABE, Protein–DNA complex, Structure

1 Introduction

Recent advances in structure determination are now yielding structures for thousands of proteins. However, obtaining the relevant conformation of a DNA-binding protein within a large multi-subunit protein–DNA complex is still challenging. For example, as of yet no complete structure of an RNA polymerase with an activator and promoter DNA is available. Even when structures are available, it is important to ascertain that the crystallographic structure is indeed the biologically relevant complex.

The chemical cleaving reagent iron bromoacetamidobenzyl-EDTA (FeBABE) is a biochemical tool that employs the Fenton reaction [1]. In this process, the iron catalyst generates hydroxyl radicals upon the addition of hydrogen peroxide and ascorbate.

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_3,

Thus, conjugating a FeBABE through a specific cysteine residue in the protein of interest will result in the cleavage of nearby polypeptide or nucleic acid bonds when the reaction is initiated. Typically, bonds within a distance of 18 Å will be cleaved [2, 3]. This technique has been successfully applied in a number of cases, such as determining the location of different regions of the primary sigma factor, σ^{70}, and an alternate sigma factor, σ^{s}, of *E. coli* RNA polymerase relative to promoter DNA [2, 4], confirming that σ^{70} Region 4 is positioned close to the −35 element of a promoter [5–7], and determining the orientation of an activator relative to its binding site [8–10]. Because the resolution of FeBABE cleavage is high, just 3–5 bp, we reasoned that this technique could be even more powerful if used to generate a fine-resolution map of a DNA-binding protein relative to its target DNA. In our example, we wanted to know the position of the bacteriophage T4 transcriptional activator MotA relative to its DNA binding site within a large complex consisting of MotA, the DNA, RNA polymerase, and the T4 co-activator AsiA. However, structures for only portions of MotA are known, and no structure of either full-length MotA or any portion of MotA together with the DNA has been obtained. Despite this, combining the FeBABE cleavage data amassed from a set of conjugated proteins with the FeBABE positioned at 11 different surface-exposed cysteine residues yielded a 3-D picture of how the protein interacts with the DNA [11]. Thus, this technique can provide a molecular view of protein–DNA interactions in the absence of a determined structure, and in the presence of a structure, the technique provides needed biochemical validation of the predicted protein–DNA contacts.

2 Materials

2.1 FeBABE Reaction

Unless otherwise indicated, solutions are made with RNase-/DNase-free water that has been deionized and autoclaved before use.

1. 22 mM FeBABE solution: Add 87.6 µL DMSO (Sigma) to 1 mg iron bromoacetamidobenzyl-EDTA (FeBABE, Dojindo Laboratories). Distribute 10 µL aliquots to Eppendorf tubes and freeze together in a black 50 mL tube at −80 °C. Thaw an aliquot right before use.
2. 0.6 % (v/v) hydrogen peroxide: 10 µL 30 % hydrogen peroxide (Sigma) diluted with 490 µL H_2O. Prepare just before use and store on ice.
3. 100 mM sodium ascorbate: 0.198 g sodium ascorbate dissolved in a final volume of 10 mL. Store as 1 mL aliquots at −20 °C. Thaw and store on ice before use.
4. TE buffer: 10 mM Tris–Cl, pH 7.9, 1 mM EDTA.

5. 100 mM thiourea: 0.076 g thiourea dissolved in a final volume of 10 mL and stored at –20 °C.
6. 1 mg/mL calf thymus DNA: Store at –20 °C.
7. TE-saturated phenol containing 0.1 % 8-hydroxyquinoline: Store at 4 °C (*see* **Note 1**).
8. 95 % ethanol: Store at –20 °C.

2.2 Components for Making Active Transcription Complex

1. Various chemicals for protein buffers: Tris base, NaCl, EDTA (ethylenediaminetetraacetic acid), EGTA (ethylene glycol tetraacetic acid), potassium phosphate monobasic, potassium phosphate dibasic, glycerol, DTT (dithiothreitol), β-mercaptoethanol, benzamidine hydrochloride, magnesium acetate, Triton X-100, RNase-free bovine serum albumin, EGTA (ethylene glycol tetraacetic acid), potassium glutamate.
2. 1 M Tris–Cl or Tris-acetate: 121 g Tris base per liter brought to the indicated pH with HCl or acetic acid, respectively.
3. 0.25 M EDTA, pH 7.0 and 7.6: 73 g EDTA per liter brought to the indicated pH with NaOH.
4. 0.20 M EGTA, pH 7.1: 76 g EGTA per liter brought to pH 7.1 with NaOH. 1 M potassium phosphate, pH 6.5: Add 1 M potassium phosphate dibasic (174.18 g/L) to a solution of 1 M potassium phosphate monobasic (136.09 g/L) until the pH is 6.5.
5. σ buffer: 50 mM Tris–Cl, pH 8, 50 mM NaCl, 50 % glycerol, 1 mM EDTA, 0.01 % Triton X-100, 0.1 mM DTT.
6. MotA FeBABE buffer: 20 mM Tris–Cl, pH 7.9, 10 % glycerol, 1 mM EDTA, 1 mM β-mercaptoethanol, 270 mM NaCl.
7. MotA GC buffer: 200 mM potassium phosphate, pH 6.5, 1 mM DTT, 1 mM benzamidine hydrochloride, 1 mM EDTA, pH 7.0, 1 mM EGTA, pH 7.1, 50 % glycerol.
8. AsiA buffer: 20 mM Tris–Cl, pH 8, 10 % glycerol, 1 mM EDTA, 1 mM β-mercaptoethanol, 50 mM NaCl.
9. Core buffer: 50 mM Tris–Cl, pH 7.5, 0.1 mM EDTA, 50 % glycerol, 250 mM NaCl, 1 mM DTT.
10. 5× KGlu transcription buffer: 40 mM Tris-acetate, pH 7.9, 150 mM potassium glutamate, 4 mM magnesium acetate, 0.1 mM EDTA, 0.1 mM DTT, 100 μg/mL bovine serum albumin.
11. 1 mg/mL heparin.
12. 10 M ammonium acetate: 7.7 g $C_2H_3O_2NH_4$ per 10 mL solution. Autoclave and store at room temperature.
13. Formamide load solution: Deionize 4 mL of formamide by mixing with AGR501-X8 (D) mixed bed resin (Bio-Rad) for at least 15 min and then filtering through a 0.22 micron syringe-driven filter unit (Millex-GP, Millipore) to remove the resin.

Combine 0.5 mL of the deionized resin with 20 μL 0.25 M EDTA, pH 7, 5 μL saturated solution of xylene cyanol FF, and 5 μL saturated solution of bromophenol blue.

14. RNaseZap: Life Technologies/Ambion.
15. Dialysis equipment: Novagen D-Tube Dialyzer Mini 6–8 kDa MWCO.

2.3 3-D Printing

3-D (three-dimensional) printed models of the proteins are produced using the ZPrinter 450 3-D printer (manufactured by Z Corporation, now 3D Systems Corp.) that is controlled by ZPrint software. Materials needed for printing are available from 3D Systems Corp.: zp151 powder, zb63 clear binder, HP11 Printhead, HP57 Color Printhead, and ColorBond infiltrant.

2.4 Molsoft Software

Visualization of the molecular interactions between the protein of interest and the DNA is performed using ICM-Pro software available through Molsoft, L.L.C., La Jolla, CA.

3 Methods

All biochemical procedures are carried out at 4 °C unless otherwise indicated and are performed under RNase-/DNase-free conditions. Disposable plasticware that has never been used or touched, such as Eppendorf tips, Eppendorf tubes, disposable cylinders, pipettes, etc., can be used directly. Otherwise, plastic and glassware are cleaned with RNaseZap and, if possible, autoclaved before use. It is important to wear clean gloves and change gloves often to minimize DNase and RNase contamination.

3.1 Conjugating FeBABE to a Protein that Contains a Single Cysteine

Dialyze 2 nmol of protein in 50 μL of its protein buffer (*see* **Note 2**) at 4 °C for at least 1 h in 1 L of an appropriate protein buffer lacking reducing agent using the D-Tube Dialyzer (*see* **Note 3**). After dialysis, transfer the protein to an Eppendorf tube. Add 0.5 μL of the FeBABE solution to the protein and incubate at 37 °C for 1 h. Repeat dialysis, but in this case, use the protein buffer containing reducing agent. (This step removes excess FeBABE.)

3.2 Forming a Stable Protein–DNA Complex

The buffer and reaction conditions for making the protein–DNA complex of interest will depend on the protein being analyzed. As an example, the conditions used for making the complex of *E. coli* RNA polymerase together with the bacteriophage T4 co-activator AsiA, T4 transcriptional activator MotA, and a 200 bp DNA fragment containing a T4 middle promoter are given here (*see* **Note 4**).

1. Incubate 0.2 μL AsiA (24 pmol in AsiA buffer) and 0.53 μL of σ^{70} (1.2 pmol in σ buffer) in an Eppendorf tube at 37 °C for 10 min. Add 1.7 pmol core RNA polymerase (1.25 μL in core

buffer), 1 μL 5× KGlu transcription buffer, and 4.02 μL H_2O and incubate for an additional 10 min at 37 °C.

2. Add a solution containing 1.0 μL of MotA (3.6 pmol in MotA buffer), 1 μL DNA that has been ^{32}P-end labeled on either the top or bottom strand (0.2–0.5 pmol in TE or H_2O; *see* **Note 5**), and 1 μL 5× KGlu transcription buffer and incubate for 10 min at 37 °C.
3. Add 0.5 μL of the heparin solution (1 mg/mL). Incubate at 37 °C for 50 s before initiating the FeBABE cleavage.

3.3 FeBABE Cleavage

1. Open the cap of the tube while it remains in the 37 °C incubator, and add 2 μL of ice-cold 100 mM sodium ascorbate to the lid and 2 μL of ice-cold 0.6 % hydrogen peroxide to a different part of the lid. Immediately close the lid and spin the tube for 1 s in a microcentrifuge, flick the tube two times to mix, and return the tube to the 37 °C incubator. (For tubes being treated with water as a control, the procedure is the same except add 4 μL of ice-cold water.)
2. After 10 min, quench the reaction by adding 80 μL 100 mM thiourea at room temperature directly to the solution in the tube. Then, add 30 μL TE (at room temperature) and 0.5 μL of ice-cold calf thymus DNA. Mix well by inverting the tube multiple times and place on ice.
3. Once all the reactions have been collected, add 125 μL TE-saturated phenol to each tube. Mix the phenol and aqueous layers well by inverting at least ten times. Place at room temperature.
4. Wait for 20 s and repeat mixing. Wait for another 20 s and repeat mixing.
5. Centrifuge tubes at 3000 × *g* for 30 s at room temperature to separate the phenol and aqueous layers. Collect tubes at room temperature.
6. Transfer the aqueous layer to another Eppendorf tube, which is on ice. Add 25 μL 10 M ammonium acetate, 500 μL ice-cold 95 % ethanol, and 0.5 μL 1 mg/mL calf thymus DNA. Mix well by inverting at least ten times.
7. Collect tubes on dry ice and let sit for at least 20 min.
8. Load tubes in a microcentrifuge so that the lid connector of each tube is in a constant position (such as directly opposite the centrifuge spindle.) Precipitate DNA by centrifuging at 14,000 × *g* for 30 min.
9. Carefully remove one tube from the microcentrifuge, and using a sterile Pasteur pipette with a long drawn-out tip, remove the supernatant by extracting the liquid from the side opposite the precipitate. Although the precipitate may not be

visible, it will be possible to know where the pellet is located by knowing how the tube was positioned within the centrifuge.

10. Place the tube on ice and repeat until all the tubes have been collected.
11. Add 200 μL ice-cold 95 % ethanol and mix thoroughly to wash the precipitate.
12. Centrifuge at 14,000 × *g* for 5 min, again loading the tubes so that the position of the pellet will be known.
13. Repeat procedure for removing the supernatant from each tube, and collect tubes on ice. Cover the lid of each tube with a ½-inch square piece of Parafilm wrapped around the top and make two or three small holes in the Parafilm using a PIPETMAN tip.
14. Dry the precipitate by vacuuming for 1 min in a SpeedVac. Dissolve the precipitate in 10 μL formamide load.
15. Electrophorese the entire volume on a 1 mm DNA sequencing gel. We use 5 % polyacrylamide (acrylamide/bis, 19:1), 7 M urea denaturing gels run in ½ X TBE for DNAs that are ~200 bp in length (*see* **Note 6**).
16. Visualize the DNA either by autoradiography or by phosphorimaging. Determine the position of each cut site by comparing the migration of the cleavage product to a Maxam–Gilbert G + A ladder also run on the gel. It is important that gels run straight and the G + A ladder is loaded in different lanes so that any curvature in the gel will not adversely affect the positioning of a cut site (Fig. 1). Preelectrophoresis of the gel for several hours before loading results in better outcomes.

3.4 Generating 3-D Models of Protein and DNA Structures

1. Using PyMOL [licenses available from Schrödinger (www.pymol.org)] or another structure analysis program, display the protein and DNA structures (*see* **Note 7**) and color-code the FeBABE-conjugated residues on the protein with their respective cut sites on the DNA. For this project, we used the structure of the C-terminal domain (CTD) of MotA (PDB ID: 1KAF; [12]), the structure of the N-terminal domain (NTD) of MotA (PDB ID: 1I1S; [13]), and double-stranded (ds) B-form DNA representing the sequence at the MotA binding site that was generated within the Molsoft program (*see* **Note 8**). Save the generated files with a ".wrl" extension.
2. Import the ".wrl" file(s) into the ZPrint program and arrange the protein model(s) as desired using the ZPrint interface (*see* **Note 9**).
3. Set the program to print. Because this will take several hours, depending on the size of the model(s), it is conveniently done overnight.

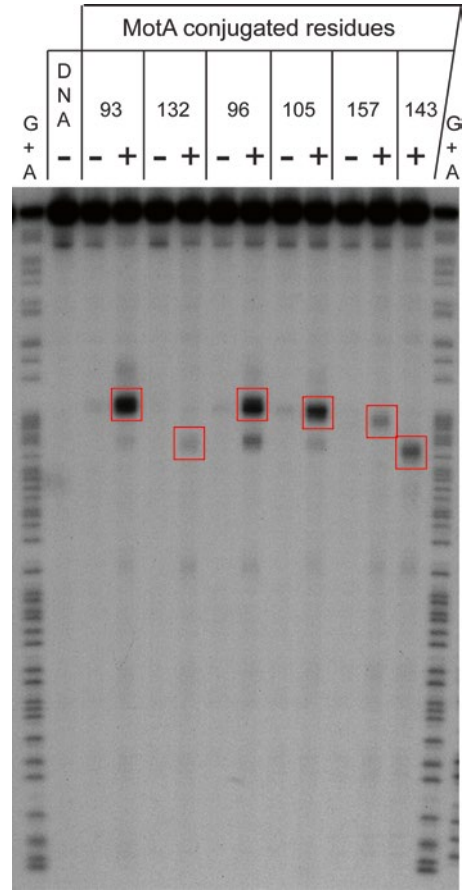


Fig. 1 DNA sequencing gel showing FeBABE-generated cleavage sites using some of the MotA proteins conjugated at the indicated residues and treated with water (–) or H_2O_2/ascorbate (+). The major cleavage sites are *boxed in red*. Positions of the cut sites are determined from the migration of these bands relative to the G + A ladder generated from the same labeled DNA

4. Brush and air-blow the models in the cleaning station to remove excess powder (Fig. 2). Handle the models carefully—they can break! Then, carefully remove the model(s) from the print chamber (*see* **Note 10**).
5. Dip the models in the ColorBond infiltrant and air-dry on wax paper. For the first few minutes, move the model(s) around on the wax paper so they do not stick.

3.5 Using 3-D Models and the Molsoft Program to Map the Protein on the DNA

1. Using the color-coded FeBABE-conjugated residues and the DNA cleavage sites, manually rotate the protein relative to the DNA to determine the most reasonable orientation for the protein relative to the DNA. In our case, an initial prediction of the best-fit manual model was found by manually moving

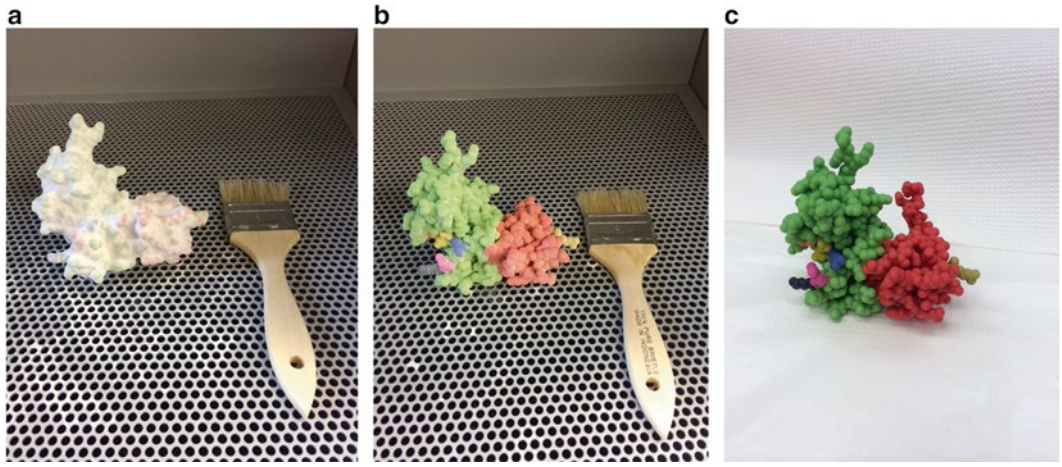


Fig. 2 A 3-D printed model of the *Bacillus subtilis* transcriptional regulator Spx (*green*) in a complex with the CTD of the α-subunit of RNA polymerase (*red*) (PDB: 3GFK; [17]), immediately after the print (**a**), after brushing to remove excess powder (**b**), and after dipping in the ColorBond infiltrant (**c**)

the MotA^{CTD} 3-D model relative to the 3-D print of the B-form DNA (Fig. 3).

2. Display the structures of the DNA and the protein domains as ICM objects in Corey–Pauling–Koltun (CPK) representation. Move the objects until they are positioned similarly to the manual model. In our case, we performed this separately for the MotA^{CTD}/DNA and MotA^{NTD}/DNA. Distance calculations between each FeBABE-conjugated residue and the DNA are then determined. The protein object is moved within the program relative to the DNA to achieve the best correlation between the predicted FeBABE cut sites and the observed cut sites. It is important to remember that FeBABE is thought to cleave within about 18 Å, so the distance between the protein residue and the DNA cleavage site should be within this distance or less. However, in many cases such as ours, distortion of the DNA caused by protein binding is not known. Consequently, for some cut sites, the distance may be greater than expected.

4 Notes

1. Phenol solution is colorless, but the addition of 8-hydroxyquinoline will turn the solution a lemon yellow color. The 8-hydroxyquinoline helps to retard oxidation of the phenol, which turns phenol red, but eventually the solution will become orange. At this point, safely discard the solution and make another.

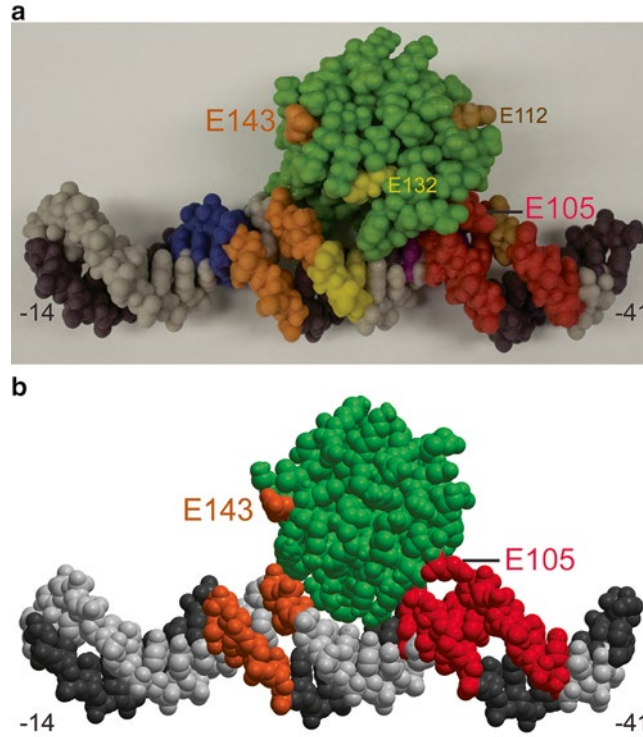


Fig. 3 Interaction of the CTD of the T4 transcriptional activator MotA with DNA [11]. (**a**) Manually determined complex using the 3-D printed structures. The FeBABE- conjugated residues within MotA^{CTD} that can be seen in this orientation (E105 [*red*], E112 [*brown*], E132 [*yellow*], E143 [*orange*]) and their respective cut sites within the DNA are color-coded. The ds DNA is shown from positions -14 to -41 relative to the start site of transcription (+1). The *top* (non-template) strand is in *black*; the *bottom* (template) strand is in *gray*. (**b**) Recreation of the complex in the ICM Molsoft program. Residues E105 (*red*) and E143 (*orange*) and their respective cleavage sites are shown

2. As FeBABE conjugates through the sulfhydryl of a cysteine, it is necessary to construct proteins containing a single cysteine at the position of interest. Start with an expression vector that contains the gene for the DNA-binding protein. Using standard cloning techniques, generate a gene in which all the natural cysteine residues have been changed to either alanines or serines. (These substitutions are considered the least harmful for cysteine.) Check that this "cys-less" version of the protein is active biochemically using the appropriate assays for the protein of interest. Using the "cys-less" gene plasmid, construct mutant genes in which specific residues are singly replaced with a cysteine. (If structural information about the protein is available, either from a structure of the protein itself (or portions of the protein) or from a structure of a homologous pro-

tein, this can be used as a guide to predict suitable surface residues for cysteine replacements.) Again check any of the generated mutant proteins for activity. It is also necessary to check the level of FeBABE conjugation. In our case, we have used mass spectrometry to determine the relative levels of conjugated and unconjugated proteins. A chemical procedure has also been reported [1].

3. Generally speaking, it is best to start with the standard buffer for the protein, but lacking reducing agent. We have found a number of typical protein buffers lacking a reducing agent to be suitable for conjugations. For example, we have successfully conjugated MotA in either MotA FeBABE buffer, which lacks a reducing agent, or MotA GC buffer minus DTT. In addition, we have conjugated σ^{70} in σ buffer minus DTT, and we have conjugated the *B. pertussis* response regulator BvgA in BvgA buffer minus DTT (20 mM HEPES, pH 7.4; 10 mM $MgCl_2$; 50 mM KCl; 50 % glycerol).
4. When working with large complexes, it is important to use conditions in which the active complex of interest is essentially the only species. In the case of a protein–DNA complex such as a transcription complex, it is often possible to employ conditions that minimize the formation of nonspecific complexes and/or to challenge unstable complexes with a specific inhibitor, such as the polyanion heparin or poly dI-dC. Thus, only stable, specific complexes will survive the competitor challenge. In our case, previous work had demonstrated that neither complexes of MotA with the DNA nor nonspecific RNAP/DNA complexes survive a 1-min challenge with heparin at 37 °C [14, 15]. Thus, we could be assured that the only species we analyzed was the active complex containing the full complement of needed proteins and DNA.
5. DNA is labeled with ^{32}P at the 5′ end of either the top or bottom strand by standard protocols. Start with oligonucleotides of ~20 nucleotides, whose sequences are approximately 100 bases upstream (for top strand labeling) or 100 bases downstream (for bottom strand labeling) of the DNA binding site. (Many commercial sources are available for purchasing oligonucleotides; we routinely use Operon.) One oligomer is treated with T4 polynucleotide kinase in the presence of ^{32}P γ-ATP to obtain a labeled primer for PCR. We use labeled ATP from PerkinElmer and the protocol and OptiKinase from USB Corporation. The labeled primer together with the unlabeled primer is then used to generate a labeled PCR product of ~200 bp. (We use PfuTurbo DNA polymerase from Stratagene for PCR.) DNA can be cleaned by a number of protocols including a PCR clean-up kit (Promega) or electroelution after electrophoresis through a native acrylamide gel using the Elutrap® electroelution system (Whatman).

6. Separate the cleaved DNA products on 7 M urea, polyacrylamide gels run in 0.5× TBE [16]. Also generate and load a Maxam–Gilbert G+A ladder [16], obtained from the same ^{32}P end-labeled DNA, on the same gel so that the positions of cut sites can be determined.
7. It is important to hide any extraneous molecules shown in the program, such as H_2O, metal ions, or ligands, so they are not printed.
8. Generation of nucleic acids is not available on the free version of Molsoft ICM. However, other programs are available for generating nucleic acid helices such as "make-na" found at http://structure.usc.edu/make-na.
9. A dialog box in ZPrint will ask for the units in which the part was created (select "inches") and the scaling because the part is too big to fit. Use the "relative" box to set the percentage, usually between 4 % and 10 %. The program will show a representation of the model in the chamber, which can be moved in 3 dimensions. If the model is too large or too small, the size can be changed. The model can also be rotated.
10. Breaks in the models can be repaired using Krazy Glue (Toagosei Co.). The glue works better if the model has not yet been dipped in the ColorBond infiltrant.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

References

1. Greiner DP, Miyake R, Moran JK, Jones AD, Negishi T, Ishihama A, Meares CF (1997) Synthesis of the protein cutting reagent iron (S)-1-(p-bromoacetamidobenzyl)ethylenediaminetetraacetate and conjugation to cysteine side chains. Bioconjugate Chem 8:44–48
2. Owens JT, Chmura AJ, Murakami K, Fujita N, Ishihama A, Meares CF (1998) Mapping the promoter DNA sites proximal to conserved regions of sigma 70 in an Escherichia coli RNA polymerase-lacUV5 open promoter complex. Biochemistry 37:7670–7675
3. Pai KS, Bussiere DE, Wang F, White SW, Bastia D (1996) Structure of the replication terminus-terminator protein complex as probed by affinity cleavage. Proc Natl Acad Sci U S A 93:10647–10652
4. Colland F, Fujita N, Kotlarz D, Bown JA, Meares CF, Ishihama A, Kolb A (1999) Positioning of sigma(S), the stationary phase sigma factor, in Escherichia coli RNA polymerase-promoter open complexes. EMBO J 18:4049–4059
5. Marr MT, Datwyler SA, Meares CF, Roberts JW (2001) Restructuring of an RNA polymerase holoenzyme elongation complex by lambdoid phage Q proteins. Proc Natl Acad Sci U S A 98:8972–8978
6. Bown JA, Owens JT, Meares CF, Fujita N, Ishihama A, Busby SJ, Minchin SD (1999) Organization of open complexes at Escherichia coli promoters. Location of promoter DNA sites close to region 2.5 of the sigma70 subunit of RNA polymerase. J Biol Chem 274:2263–2270

7. Marr MT, Roberts JW, Brown SE, Klee M, Gussin GN (2004) Interactions among CII protein, RNA polymerase and the lambda PRE promoter: contacts between RNA polymerase and the -35 region of PRE are identical in the presence and absence of CII protein. Nucleic Acids Res 32:1083–1090
8. Boucher PE, Maris AE, Yang MS, Stibitz S (2003) The response regulator BvgA and RNA polymerase alpha subunit C-terminal domain bind simultaneously to different faces of the same segment of promoter DNA. Mol Cell 11:163–173
9. Decker KB, Chen Q, Hsieh ML, Boucher P, Stibitz S, Hinton DM (2011) Different requirements for sigma Region 4 in BvgA activation of the Bordetella pertussis promoters Pfim3 and PfhaB. J Mol Biol 409:692–709
10. King JM, Brutinel ED, Marsden AE, Schubot FD, Yahr TL (2012) Orientation of Pseudomonas aeruginosa ExsA monomers bound to promoter DNA and base-specific contacts with the P(exoT) promoter. J Bacteriol 194:2573–2585
11. Hsieh ML, James TD, Knipling L, Waddell MB, White S, Hinton DM (2013) Architecture of the bacteriophage T4 activator MotA/promoter DNA interaction during sigma appropriation. J Biol Chem 288:27607–27618
12. Li N, Sickmier EA, Zhang R, Joachimiak A, White SW (2002) The MotA transcription factor from bacteriophage T4 contains a novel DNA-binding domain: the 'double wing' motif. Mol Microbiol 43:1079–1088
13. Li N, Zhang W, White SW, Kriwacki RW (2001) Solution structure of the transcriptional activation domain of the bacteriophage T4 protein, MotA. Biochemistry 40:4293–4302
14. March-Amegadzie R, Hinton DM (1995) The bacteriophage T4 middle promoter PuvsX: analysis of regions important for binding of the T4 transcriptional activator MotA and for activation of transcription. Mol Microbiol 15:649–660
15. Hinton DM, March-Amegadzie R, Gerber JS, Sharma M (1996) Characterization of pretranscription complexes made at a bacteriophage T4 middle promoter: involvement of the T4 MotA activator and the T4 AsiA protein, a sigma 70 binding protein, in the formation of the open complex. J Mol Biol 256:235–248
16. Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65:499–560
17. Lamour V, Westblade LF, Campbell EA, Darst SA (2009) Crystal structure of the in vivo-assembled Bacillus subtilis Spx/RNA polymerase alpha subunit C-terminal domain complex. J Struct Biol 168:352–356

Chapter 4

In Cellulo DNA Analysis: LMPCR Footprinting

Régen Drouin, Nathalie Bastien, Jean-François Millau, François Vigneault, and Isabelle Paradis

Abstract

The in cellulo analysis of protein-DNA interactions and chromatin structure is very important to better understand the mechanisms involved in the regulation of gene expression. The nuclease-hypersensitive sites and sequences bound by transcription factors often correspond to genetic regulatory elements. Using the ligation-mediated polymerase chain reaction (LMPCR) technology, it is possible to precisely analyze these DNA sequences to demonstrate the existence of DNA-protein interactions or unusual DNA structures directly in living cells. Indeed, the ideal chromatin substrate is, of course, found inside intact cells. LMPCR, a genomic sequencing technique that map DNA single-strand breaks at the sequence level of resolution, is the method of choice for in cellulo footprinting and DNA structure studies because it can be used to investigate complex animal genomes, including human. The detailed conventional and automated LMPCR protocols are presented in this chapter.

Key words Footprints, Ligation-mediated polymerase chain reaction, Polymerase chain reaction, DNA polymerase, Living cell, Deoxyribonuclease I, Dimethylsulfate, Ultraviolet light C, DNA-protein interaction

1 Introduction

The in cellulo analysis of protein-DNA interactions and chromatin structure can provide several critical information regarding regulation of gene expression. For example, DNA sequences spanned by nuclease-hypersensitive sites or bound by transcription factors often correspond to genetic regulatory elements. Using the ligation-mediated polymerase chain reaction (LMPCR) technology, it is possible to map DNA sequences to demonstrate the existence of DNA-protein interactions or unusual DNA structures directly in living cells. LMPCR analyses can thus be used as a primary investigative tool to identify the regulatory sequences involved in gene expression. Once specific promoter sequence sites shown to be bound by transcription factors in living cells, it is often possible to establish the identity of these factors simply by comparison with

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_4,

the consensus binding sites of known factors such as SP1, AP-1, NF-1, and so forth. The identity of each factor can then be confirmed by performing the well-established chromatin immunoprecipitation (ChIP) technique, or by using silencing RNA directed against the protein of interest.

The native state of a gene and most of the special DNA structures are unavoidably lost when DNA is cloned or purified [1–4]. Hence, the commonly used in vitro (purified DNA) footprinting and electrophoretic mobility shift assay (EMSA) cannot demonstrate that a given DNA-protein interaction occurs within the cells of interest. It is clear that gene promoters are best studied in their natural state in the living cell. Thus, it is not surprising that in cellulo (living cell) DNA footprinting is one of the most accurate predictors of the state of gene transcriptional activities [1, 3, 5]. LMPCR is the method of choice for in cellulo footprinting and DNA structure studies because it can be used to investigate complex animal genomes, including human. However, the quality and usefulness of the information obtained from any in cellulo DNA analysis depends on three parameters: (1) the integrity of the native chromatin substrate used in the experiment, (2) the structural specificity of the chromatin probe, and (3) the sensitivity of the assay. The ideal chromatin substrate is, of course, found inside intact cells. However, a near-ideal chromatin substrate can still be found in permeabilized cells, allowing the application of a wider range of DNA cleavage agents, including DNase I.

In cellulo footprinting assesses the local reactivity of probing agents on living cells DNA as compared to that on purified DNA (*see* Figs. 1, 2, 3, and 4). Two steps are necessary for in cellulo footprinting analysis: (1) the treatment of purified DNA and intact cells with a given probing agent followed, if necessary, by DNA damage conversion into single-strand breaks and (2) the analysis of amplified DNA fragments on a sequencing gel. A comparison is then made between the modification frequency obtained in vitro and in cellulo. For example, each guanine residue of purified DNA has a near-equivalent probability of being methylated by dimethylsulfate (DMS) and thus, the cleavage pattern of in vitro modified DNA appears on a sequencing gel as a ladder of bands of roughly equal intensity. However, in the presence of DNA-binding proteins (transcription factors), all guanine residues do not show the same accessibility to DMS in living cells (Fig. 1). Thus, differences between banding patterns obtained from in vitro- and in cellulo-modified DNA can be used to infer protein-binding sites in living cells. It is always advisable to validate such interpretations using more than one footprinting agent.

The measure of in cellulo footprints has historically been problematic due to the dilute nature of target sequences and the complexity of higher eukaryotes genomes. The development of LMPCR, an extremely sensitive and specific technique, resolved this problem. The LMPCR technique quantitatively maps DNA

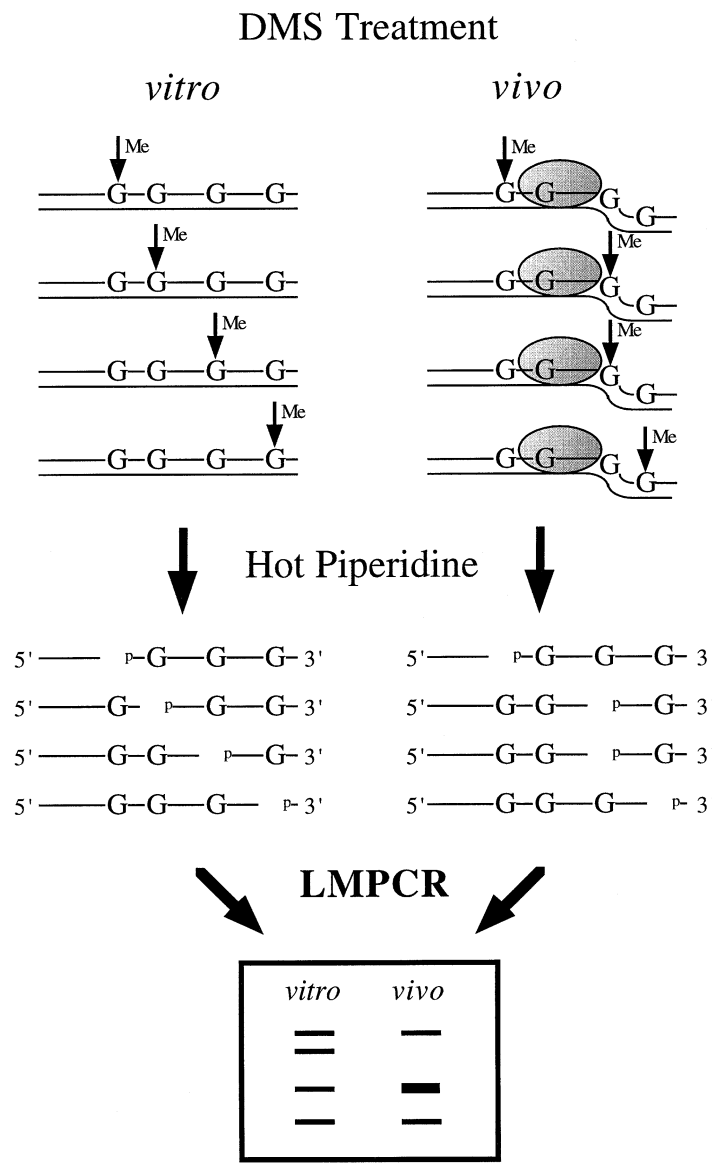


Fig. 1 Overall scheme for in vivo DNA analysis using DMS. The methylation of guanine residues following DMS treatment of purified DNA (in vitro) and cells (in vivo) is shown with *vertical arrows* and methylated residues (Me). When purified DNA is treated with DMS, every guanine residue has a similar probability of being methylated. However, the guanine residue in intimate contact with a sequence-specific DNA-binding protein illustrated by the *dotted oval* is protected from DMS methylation, whereas the guanine residue localized close to the boundary of the DNA-protein contact that modifies DNA structure, allowing a better accessibility to DMS, is methylated more frequently. The methylated guanine residues are cleaved by hot piperidine leaving phosphorylated 5′ ends. On the sequencing ladder following LMPCR, guanine residues that are protected from methylation appear as missing or less intense bands when compared with the sequencing ladder from the same DNA sequence obtained after DMS treatment of purified DNA. On the other hand, guanine residues that undergo enhanced DMS methylation appear as darker bands in the sequencing ladder relative to the purified DNA control

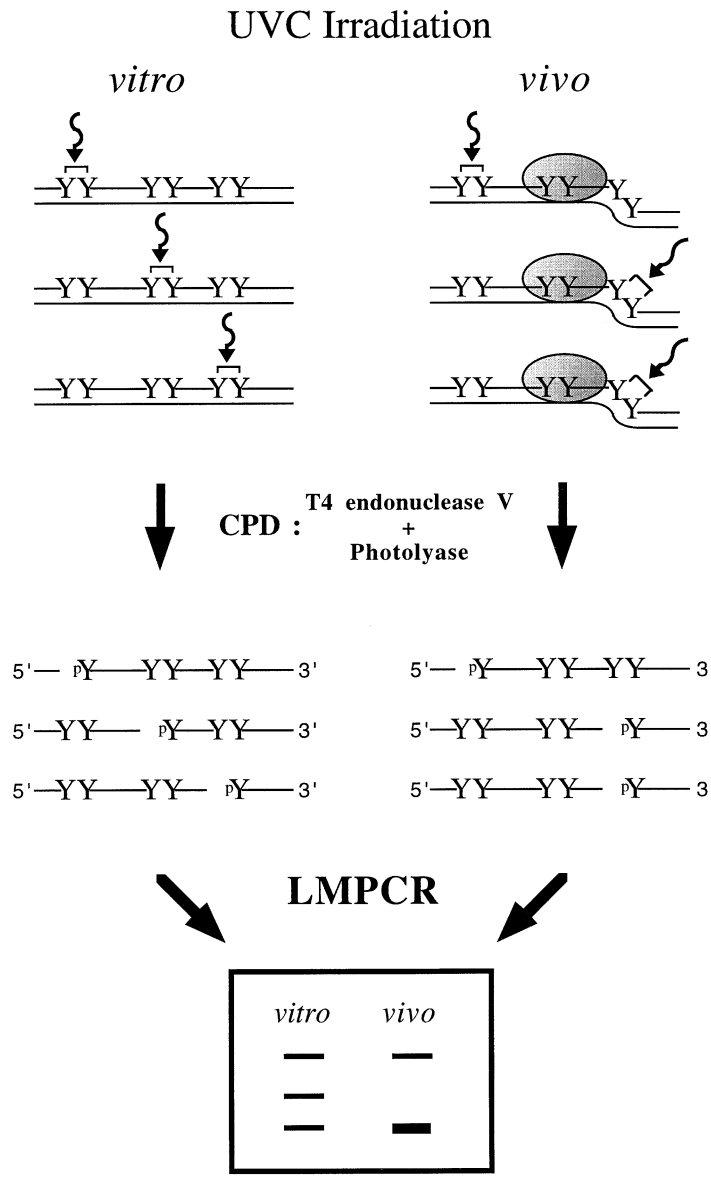


Fig. 2 Overall scheme for in vivo DNA analysis using UVC and CPD formation. The CPD formation following UVC exposure of purified DNA (in vitro) and cells (in vivo) is shown with *curved arrows* and *brackets* linking two adjacent pyrimidines (Y). When purified DNA is irradiated with UVC, the frequency of CPD formation at dipyrimidine sites is determined by the DNA sequence. However, the presence of a sequence-specific DNA-binding protein illustrated by the dotted oval as well as DNA structure can prevent (negative photofootprint) or enhance (positive photofootprint) CPD formation. The CPD is cleaved by T_4 endonuclease V digestion and photolyase photoreactivation leaving phosphorylated 5′ ends. On the sequencing ladder following LMPCR, the negative photofootprints appear as missing or less intense bands when compared with the sequencing ladder from the same DNA sequence obtained after UVC irradiation of purified DNA. On the other hand, positive photofootprints appear as darker bands in the sequencing ladder relative to the purified DNA control

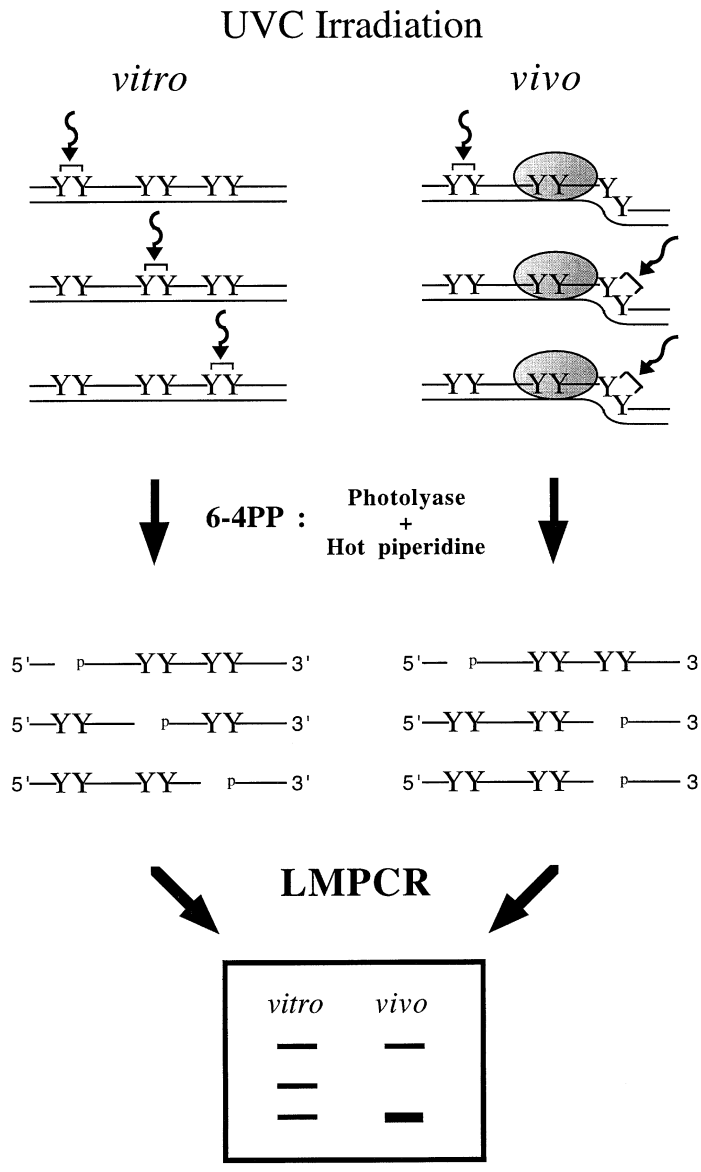


Fig. 3 Overall scheme for in vivo DNA analysis using UVC and 6-4PP formation. The 6-4PP formation following UVC exposure of purified DNA (in vitro) and cells (in vivo) is shown with *curved arrows* and *brackets* linking two adjacent pyrimidines (Y). When purified DNA is irradiated with UVC, the frequency of 6-4PP formations at dipyrimidine sites is determined by the DNA sequence. However, the presence of a sequence-specific DNA-binding protein illustrated by the dotted oval as well as DNA structure can prevent (negative photofootprint) or enhance (positive photofootprint) 6-4PP formation. First, CPD are photoreactivated by photolyase and then 6-4PP are cleaved by hot piperidine treatment leaving phosphorylated 5′ ends. On the sequencing ladder following LMPCR, the negative photofootprints appear as missing or less intense bands when compared with the sequencing ladder from the same DNA sequence obtained after UVC irradiation of purified DNA. On the other hand, positive photofootprints appear as darker bands in the sequencing ladder relative to the purified DNA control

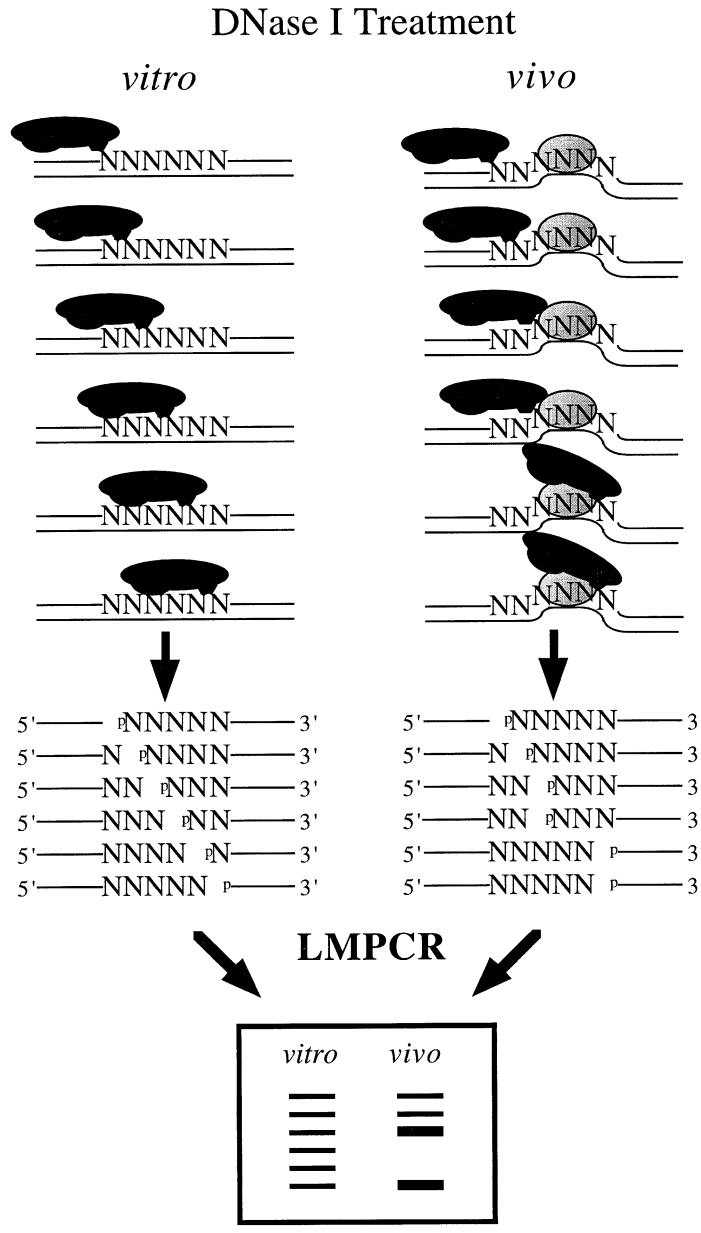


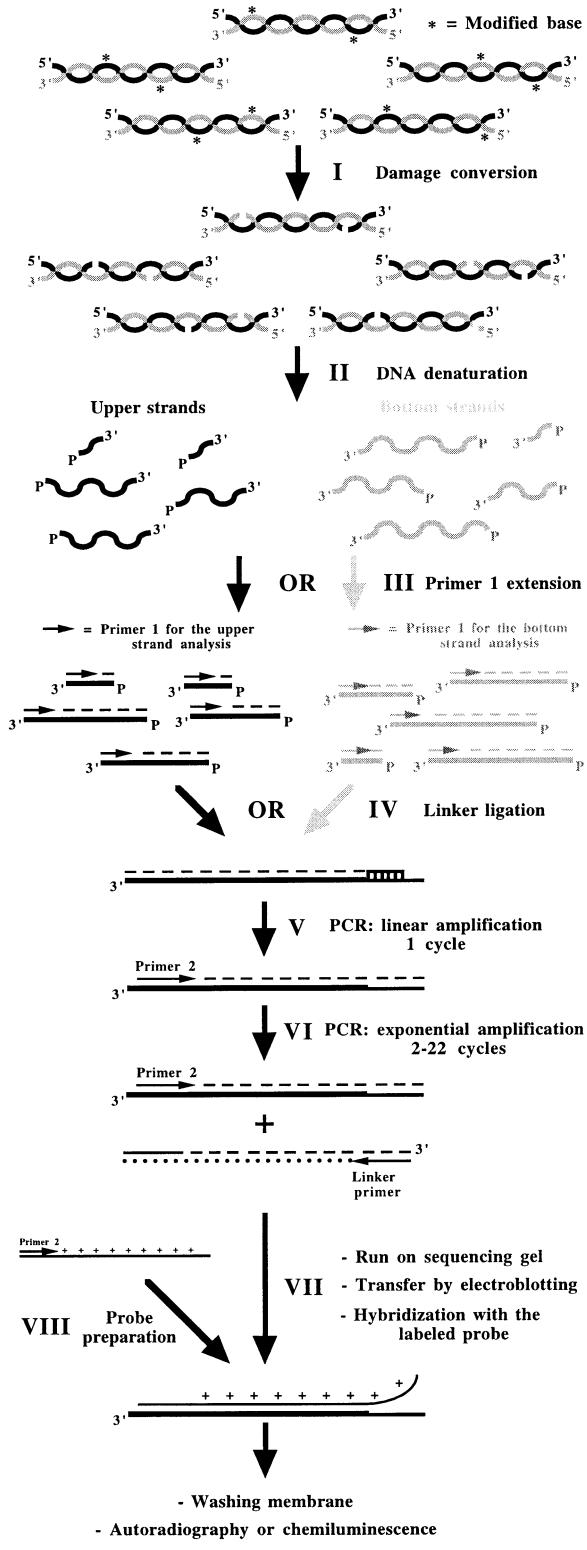
Fig. 4 Overall scheme for in vivo DNA analysis using DNase I. The DNase I enzyme (the solid black) digestion of purified DNA (in vitro) and cells (in vivo) is shown. When purified DNA is digested with DNase I, the cleavage pattern shows that sites of the nucleotide sequence have similar probabilities of being cleaved. However, the presence of a sequence-specific DNA-binding protein illustrated by the dotted oval as well as DNA structure can prevent (protection) or enhance (hypersensitive) DNase I cleavage. The DNase I cleavage leaves phosphorylated 5′ ends. On the sequencing ladder following LMPCR, DNA sequences that are protected from DNase I cleavage appear as missing or less intense bands when compared with the sequencing ladder from the same DNA sequence obtained after DNase I digestion of purified DNA. On the other hand, hypersensitive sites that undergo enhanced DNase I cleavage appear as darker bands in the sequencing ladder relative to the purified DNA control

single-strand breaks having phosphorylated 5′ ends within single-copy DNA sequences. It was first developed by Mueller and Wold [6] for DMS footprinting and subsequently, Pfeifer and colleagues adapted it to DNA sequencing [7], methylation analyses [1, 7], DNase I footprinting [5], nucleosome positioning [5], and UV footprinting (photofootprinting) [4, 8]. LMPCR can be combined with a variety of DNA-modifying agents used to probe the chromatin in cellulo. No single technique can provide as much information on DNA-protein interactions and DNA structures existing within the living cells as LMPCR can.

1.1 General Overview of LMPCR

Genomic sequencing techniques, such as developed by Church and Gilbert [9] can be used to map strand breaks in mammalian genes at the nucleotide resolution. However, by incorporating an exponential amplification step, LMPCR (outlined in Fig. 5) is advantageously more sensitive. It uses 20 times less DNA to obtain a nucleotide-resolution banding pattern and allows shorter autoradiographic exposure times than this technique. The unique aspect of LMPCR is the blunt-end ligation of an asymmetric double-stranded linker (5′ overhanging to avoid self-ligation or ligation in the wrong direction) onto the 5′ end of each cleaved blunt-ended DNA molecule [6, 7]. The blunt end is created by the primer extension (PE) of a gene-specific primer (primer 1 in Fig. 5) until a strand break is reached. Because the generated breaks are randomly distributed along the genomic DNA and thus have 5′ ends of unknown sequence, the asymmetric linker adds a common and known sequence to all 5′ ends. This then allows exponential PCR amplification using the longer oligonucleotide of the linker (linker-primer) and a second nested gene-specific primer (primer 2 in Fig. 5). LMPCR preserves the quantitative representation of each fragment in the original population of cleaved molecules [10–13]. After 22 cycles of PCR, the DNA fragments are size-fractionated on a sequencing gel, allowing quantification on a phosphorimager [14–17]. Thus, the band intensity pattern obtained by LMPCR directly reflects the frequency distribution of 5′-phosphoryl DNA breaks along a 200 bp (or more if a sequencer is used) sequence adjacent to the nested primer.

Three variations of the LMPCR technique do exist. Pfeifer and colleagues [7] took advantage of electroblotting DNA onto a nylon membrane followed by hybridization with a gene-specific probe to reveal the sequence ladders. This probe is typically ^{32}P-radiolabeled, but some have successfully used digoxigenin to get ride of the radioactivity issue. (See detailed protocol in the second edition of *Methods in molecular biology*, [18].) On the other hand, Mueller and Wold [6] used a nested third radiolabeled primer for the last one or two cycles of the PCR amplification step. It is worthwhile to note that this last technique was recently employed to analyze LMPCR-amplified DNA fragments using


* = Modified base
I Damage conversion
II DNA denaturation
Upper strands
Bottom strands
OR
III Primer 1 extension
= Primer 1 for the upper strand analysis
= Primer 1 for the bottom strand analysis
OR
IV Linker ligation
V PCR: linear amplification 1 cycle
Primer 2
VI PCR: exponential amplification 2-22 cycles
Primer 2
Linker primer
Primer 2
VII - Run on sequencing gel
- Transfer by electroblotting
- Hybridization with the labeled probe
VIII Probe preparation
- Washing membrane
- Autoradiography or chemiluminescence

sequencer devices [19, 20]. In this chapter, we describe two LMPCR protocols routinely used in the laboratory: one derived from the Pfeifer and colleagues protocol and the other based on the Mueller and Wold alternative sequencer method.

We also describe three probing methods generally combined with LMPCR to reveal in cellulo protein-DNA interactions: DMS, ultraviolet (UV), and DNase I (Figs. 1, 2, 3, and 4, Table 1). These modifying agents provide complementary information and each has its associated advantages and drawbacks (Table 2). To best characterize protein-DNA interactions, it is often necessary to use two or even all three of these methods. Treatments with any probing agent must produce either strand breaks or modified nucleotides that can be converted to DNA strand breaks with a 5′-phosphate (*see* Figs. 1, 2, 3, and 4, Table 3). These protocols may also be

Table 1
Purposes of the three main in cellulo footprinting approaches

Approaches	Goals
1. Dimethylsulfate (DMS)	1. Localizes in cellulo DNA-protein contacts located in the major groove of the DNA double helix 2. Can detect special DNA structures
2. UV irradiation (UVB or UVC)	1. Localizes in cellulo DNA-protein interactions and shows how DNA structure is affected in the presence of transcription factors 2. Can detect special DNA structures 3. Can show evidence of positioned nucleosomes
3. DNase I	1. Localizes in cellulo DNA-protein contacts 2. Precisely maps in cellulo DNase I hypersensitive sites 3. Shows evidence of nucleosomes and their positions; can differentiate core DNA from linker DNA

Fig. 5 Outline of the LMPCR procedure. Step I: Specific conversion of modified bases to phosphorylated single-strand breaks. Step II: Denaturation of genomic DNA. Step III: Annealing and extension of primer 1 (although both strands can be studied; each LMPCR protocol only involves the analysis of either the non-transcribed strand or the transcribed strand). Step IV: Ligation of the linker. Step V: First cycle of PCR amplification; this cycle is a linear amplification because only the gene-specific primer 2 can anneal. Step VI: Cycle 2–22 of exponential PCR amplification of gene-specific fragments with primer 2 and the linker primer (the longer oligonucleotide of the linker). Step VII: Separation of the DNA fragments on a sequencing gel, transfer of the sequence ladder to a nylon membrane by electroblotting, and visualization of the sequence ladder by hybridization with a labeled single-stranded probe. Step VIII: Preparation and isotopic or non-isotopic labeling of single-stranded probe

Table 2
Advantages and drawbacks of the three main in cellulo footprinting approaches

Approaches	Advantages	Drawbacks
DMS	Treatment is technically easy to carry out; the DMS is a small molecule that penetrates very easily into living cells with little disruption.	1. Requires guanines, therefore is sequence dependent. 2. Does not detect all DNA-protein interactions.
UV irradiation (UVB or UVC)	1. Treatment is technically easy to carry out; UV light penetrates through the outer membrane of living cells without disruption. 2. Detects many DNA-protein interactions. 3. Very sensitive to particular DNA structures.	1. Requires two adjacent pyrimidines, therefore is sequence dependent. 2. The interpretation of the results is sometime difficult; to differentiate between DNA-protein interactions and special DNA structures can be very difficult.
DNase I	1. Little sequence dependency. 2. No conversion of modified bases required. 3. Detects all DNA-protein contacts. 4. Very sensitive to particular DNA structures.	1. Technically difficult to carry out; reproducibility is often a problem. 2. DNase I is a protein which can penetrate in living cells only following membrane permeabilization, thus causing some cell disruption.

Table 3
Mapping schemes used with the three main in cellulo footprinting approaches

Approaches	Strand Breaks	Modified bases	Conversion of modified bases to DNA single-strand breaks
DMS	Few	Guanine: methylated guanines at N7 position Adenine: to a much lesser extent, methylated adenines at N3 position	Hot piperidine
UV irradiation (UVB or UVC)	Very few	1. Cyclobutane pyrimidine dimers 2. 6–4 photoproducts	1. T4 endonuclease V followed by photolyase 2. Photolyase followed by hot piperidine
DNase I	Yes	None	No conversion is required

adapted to footprinting with other probing agents, such as $KMnO_4$ and OsO_4 [56]. However a detailed description is beyond the scope of the present chapter.

1.2 In Cellulo Dimethylsulfate Footprint Analysis (Fig. 1)

DMS is a small highly reactive molecule that easily diffuses through the outer cell membrane and into the nucleus. It preferentially methylates the N7 position of guanine residues via the major groove and, to a lesser extent, the N3 position of adenine residues

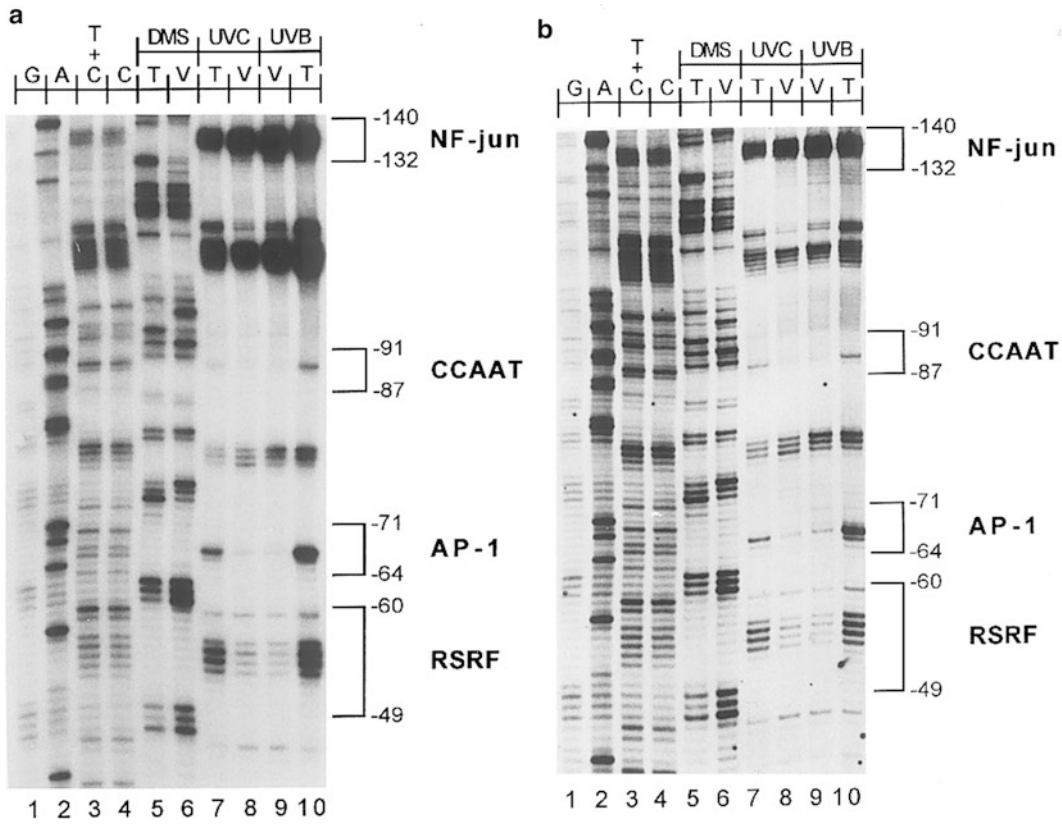


Fig. 6 LMPCR analysis of methylated guanines and CPD along the nontranscribed strand of the c-jun promoter following DMS treatment, and UVB and UVC irradiation, respectively. The membrane was hybridized with an isotopic [^{32}P]-dCTP-labeled probe. The membrane was exposed on film between two intensifying screens for 25 min at −70 °C. *Lanes 1–4*: LMPCR of DNA treated with chemical cleavage reactions. These lanes represent the sequence of the *c-jun* promoter analyzed with JD primer set [39]. *Lanes 5–6*: LMPCR of DMS-treated naked DNA (T: in vitro) and fibroblasts (V: in vivo) followed by hot piperidine treatment. *Lanes 7–10*: LMPCR of UVC- and UVB-irradiated naked DNA (T) and fibroblasts (V) followed by T_4 endonuclease V/photolyase digestions. On the right, the consensus sequences of transcription factor binding sites are delimited by *brackets*. The numbers indicate their positions relative to the major transcription initiation site

via the minor groove. The most significant technical advantage of in cellulo DMS footprinting is that DMS can be simply added to the cell culture medium (*see* Table 2 for advantages and drawbacks). Each guanine residue of in vitro DNA displays the same probability of being methylated by DMS. Because DNA inside living cells forms chromatin and is often found associated with a number of proteins, it is expected that its reactivity toward DMS will differ from that of in vitro DNA. Figures 6 and 7 show in cellulo DMS treatment patterns compared to the treatment of in vitro DNA. Proteins in contact with DNA either decrease accessibility of specific guanines to DMS (protection) or, as frequently observed at the edges of a footprint, increase reactivity (hyperreactivity) [1].

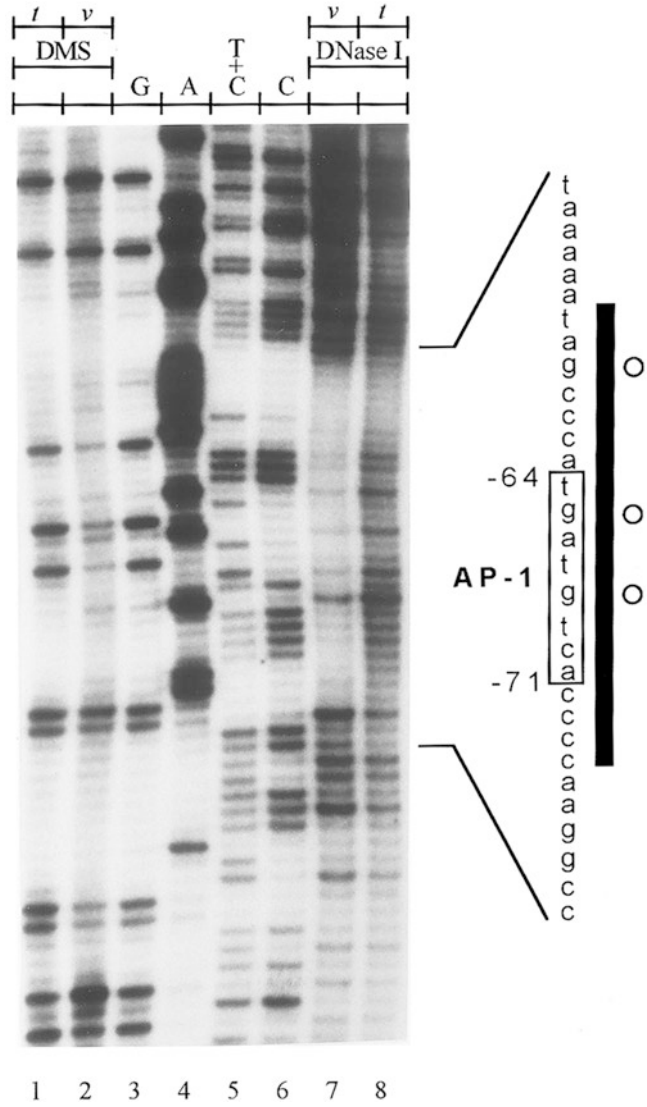


Fig. 7 LMPCR analysis of methylated guanines and DNA strand breaks along the transcribed strand of the c-jun promoter following DMS treatment and DNase I digestion, respectively. The membrane was hybridized with an isotopic [^{32}P]-dCTP-labeled probe. *Lanes 1–2*: LMPCR of DMS-treated purified DNA (t: in vitro) and fibroblasts (v: in vivo) followed by hot piperidine treatment. *Lanes 3–6*: LMPCR of DNA treated with chemical cleavage reactions. These lanes represent the sequence of the *c-jun* promoter analyzed with JC primer set [39]. *Lanes 7–8*: LMPCR of DNase I-digested permeabilized fibroblasts (v) and purified DNA (t). As a reference, a small portion of the chemically derived sequence is shown on the right of the autoradiogram, the AP-1-like binding sequence is enclosed by a *box*, and the numbers indicate its position relative to the major transcription initiation site. *Open circles* represent guanines that are protected against DMS-induced methylation (negative DMS footprints) in vivo. The *black bar* shows the protected sequence against DNase I-induced cleavage in vivo. Thus, in vivo DNase I footprinting analysis delimits much better the DNA-protein interactions

Hyperreactivity can also indicate a greater DMS accessibility of special in cellulo DNA structure [21]. Hot piperidine cleaves the glycosylic bond of methylated guanines and adenines, leaving a ligatable 5′-phosphate [22].

Genomic footprinting using DMS reveals DNA-protein contacts located in the major groove of the DNA double helix (Table 1). However, it should be noted that in cellulo DNA analysis studies using DMS alone may not detect some DNA-protein interactions [23]. First, no DNA-protein interactions are detected in the absence of guanine residues. Second, some proteins do not affect DNA accessibility to DMS. Third, certain weak DNA-protein contacts could be disrupted because of the high reactivity of DMS. Thus, when using DMS, it is often important to confirm results with alternative in cellulo footprinting approaches [23, 24].

1.3 In Cellulo UV Footprint (Photofootprint) Analysis (Figs. 2 and 3)

UVC (200–280 nm) and UVB (280–320 nm) can also be used as probing agents for in cellulo footprinting [4, 8, 25–27]. When cells are subjected to UVC (254 nm) or UVB, two major classes of lesions are introduced into DNA at dipyrimidine sites (CT, TT, TC, and CC): cyclobutane pyrimidine dimers (CPD) and the pyrimidine (6-4) pyrimidone photoproducts (6-4PP) [28]. CPD are formed between the 5,6 bonds of any two adjacent pyrimidines, whereas a stable bond between positions 6 and 4 of two adjacent pyrimidines characterizes 6-4PP. 6-4PP are formed at a rate 15–30 % of that of CPD [29] and are largely converted to their Dewar valence isomers by direct secondary photolysis (photoisomerization) [29]. In living cells, the photoproduct distribution is determined both by sequence context and chromatin structure [30]. In general, CPD and 6-4PP appear to form preferentially in longer pyrimidine runs. Because DNA absorbs directly UVB and UVC and because cells are exposed during a short period of time to high UV intensities, there are relatively few perturbations of other cellular processes and secondary events that could modify the chromatin structure or release DNA-protein interactions. Thus, UV irradiation is probably one of the least disruptive footprinting methods and hence truly reflects the in cellulo situation (Table 2). As for DMS, DNA-binding proteins influence the distribution of UV photoproducts in a significant way [25]. When the photoproduct spectrum of in vitro irradiated DNA is compared with that obtained after irradiation in cellulo, differences become apparent. The photoproduct frequency within sequences bound by sequence-specific DNA-binding proteins is suppressed or enhanced in comparison to in vitro DNA [2, 4, 8]. Effects of chromatin structure may be significant in regulatory gene regions that bind transcription factors (Fig. 6). Mapping of CPD at the single-copy gene level can reveal positioned nucleosomes because CPD is modulated in a 10-bp periodicity within nucleosome core DNA [31, 32]. 6-4PP forms more frequently in linker DNA than in core DNA [33].

Photofootprints reveal variations in DNA structure associated with the presence of transcription factors or other proteins bound to DNA. UV light has the potential to reveal all DNA-protein interactions provided when a dipyrimidine sequence on either DNA strand within a putative binding sequence. Because UV footprints can be seen outside protein-binding sites, UV light should not be used as the only in cellulo footprinting agent. The precise delimitations of the protein-DNA contact are difficult to determine with the simple in cellulo UV probing method.

The distribution of UV-induced CPD and 6-4PP along genomic DNA can be mapped at the sequence level by LMPCR following conversion of these photoproducts into ligatable 5′-phosphorylated single-strand breaks. CPD are enzymatically converted by cleavage with T_4 endonuclease V followed by UVA (320–400 nm) photoreactivation of the overhanging pyrimidine using photolyase (Fig. 2) [8]. Because the 6-4PP and their Dewar isomers are hot alkali-labile sites, they can be cleaved by hot piperidine (Fig. 3) [2]. Generally we simply measure the CPD distribution. Performing 6-4PP mapping is of interest only if no other alternative footprinting method is available.

1.4 In Cellulo DNase I Footprint Analysis (Fig. 4)

DNase I treatment of permeabilized cells gives clear footprints when the DNase I-induced breaks are mapped by LMPCR [5]. As with DMS and UV, footprint analyses are obtained by comparing in cellulo DNase I digestion patterns to patterns obtained from the in vitro DNA digestion (Fig. 7). When compared to in vitro DNA, permeabilized cells show protected bands at protein-DNA interaction sequences and DNase I hypersensitive bands in regions of higher order nucleoprotein structure [5]. Compared to DMS, DNase I is less base selective, more efficient at detecting minor groove DNA-protein contacts, provides more information on chromatin structure, displays larger and clearer footprints, and better delimits the boundaries of DNA-protein interactions (Fig. 7). The nucleotides covered by a protein are almost completely protected on both strands from DNase I nicking, allowing a better delimitation of the boundaries of DNA-protein contacts. However, it should be underlined that the relatively bulky DNase I molecule cannot cleave the DNA in the immediate vicinity of a bound protein because of steric hindrance. Consequently, the regions protected from cutting can extend beyond the real DNA-protein contact site. On the other hand, when DNA is wrapped around a nucleosome-size particle, DNase I cutting activity is increased at 10-bp intervals and usually no footprint is observed (Tables 1 and 2).

The DNase I is a relatively large 31 kD protein and cannot penetrate cells without previous cell-membrane permeabilization. Cells can be efficiently permeabilized by l-α-lysophosphatidylcholine (lysolecithin) [5] or Nonidet P40 [34] (see detailed protocol on the second edition of *Methods in molecular biology*, [18]). It has

been shown that cells permeabilized by lysolecithin remain intact, replicate their DNA very efficiently and show normal transcriptional activities [35, 36]. There are numerous studies showing that lysolecithin-permeabilized cells maintain a normal nuclear structure to a greater extent than isolated nuclei, because the chromatin structure can be significantly altered during the nuclear isolation procedures [5]. Indeed, DNase I footprinting studies using isolated nuclei can be flawed because transcription factors are lost during the isolation of nuclei in polyamine containing buffers [5]. Even though other buffers may be less disruptive, transcription factors can still be lost during the isolation procedure, leading to the complete or partial loss of footprints.

DNase I digestion of DNA leaves ligatable 5′-phosphorylated breaks, but the 3′-ends are free hydroxyl groups. Pfeifer and colleagues [5, 37] observed that these 3′-OH ends can be used as short primers and extended by the DNA polymerases during the PE and/or PCR steps of LMPCR, thereby reducing significantly the overall efficiency of LMPCR and giving a background smear on sequencing gels. To avoid the nonspecific priming of these 3′-OH ends, three alternative solutions have been applied: (1) blocking these ends by the addition of a dideoxynucleotide [5, 37], (2) enrichment of fragments of interest by extension product capture using biotinylated gene-specific primers and magnetic streptavidin-coated beads [38–41], and (3) performing primer 1 hybridization and PE at a higher temperature (52–60 °C vs. 48 °C, and 75 °C vs. 48 °C, respectively, using a thermostable DNA polymerase such as *Vent* exo- and *Pfu* exo- [3, 42–45]. Although effective, the first two alternatives involve additional manipulations that are time consuming. Because of its simplicity, we select primer 1 with higher T_m (52–60 °C) and use the *Pfu* exo- or *Vent* exo- DNA polymerase for the PE.

1.5 Choice of DNA Polymerases for LMPCR

LMPCR involves the PCR amplification of differently sized genomic DNA fragments. During the LMPCR procedure, DNA polymerases are required for two steps: PE and PCR amplification. For the PE step, the best DNA polymerase would be one that (1) is thermostable and very efficient, (2) has no terminal transferase activity, (3) is able to efficiently polymerize about 0.5 kb of DNA even when the DNA is very GC rich and (4) is able to polymerize through any DNA secondary structures. For the PCR step, the best DNA polymerase would be (1) thermostable, (2) very efficient, (3) able to amplify indiscriminately a mixture of DNA fragments of different lengths (between 50 and 500 bp) and of varying GC richness (from 5 to 95 %), and (4) able to efficiently resolve DNA secondary structures. We find that *Pfu* exo- and *Vent* exo- are the best enzymes for the PE and PCR steps of LMPCR [44, 46]. In this chapter, LMPCR protocols using *Pfu* exo- DNA polymerase for PE and PCR steps are described in detail. However, because the *Vent* exo- polymerase

is frequently used for the PE step and *Taq* DNA polymerase or *Vent* exo- polymerase for the PCR amplification, an alternative protocol using these polymerases is also included.

2 Materials

Nanopure H_2O should be used in making any buffers, solutions, and dilutions, unless otherwise specified.

2.1 DNA Purification (for 10^7 to 10^8 Cells)

1. Any type of cells (i.e., fibroblasts, lymphocytes).
2. Trypsin-EDTA (Wisent).
3. Buffer A: 300 m*M* sucrose, 60 m*M* KCl, 15 m*M* NaCl, 60 m*M* Tris–HCl pH 8.0, 0.5 m*M* spermidine, 0.15 m*M* spermine, and 2 m*M* EDTA. Store at –20 °C.
4. Buffer A + 1 % Nonidet P40 Substitute (Fluka). Store at –20 °C.
5. Buffer B: 150 m*M* NaCl and 5 m*M* EDTA pH 7.8.
6. Buffer C: 20 m*M* Tris–HCl pH 8, 20 m*M* NaCl, 20 m*M* EDTA, and 1 % SDS.
7. Proteinase K from *Tritirachium album* (Roche Applied Science).
8. RNase A from bovine pancreas (Roche Applied Science).
9. Phenol, equilibrated, pH 8 (USB Corporation).
10. Chloroform.
11. 5 *M* NaCl.
12. Precooled absolute ethanol (–20 °C).
13. Precooled 80 % ethanol (–20 °C).

2.2 Chemical Cleavage for DNA Sequencing Products

1. K_2PdCl_4 solution: 10 m*M* K_2PdCl_4 (potassium tetrachloropalladate(II), Aldrich) and 100 m*M* HCl, pH 2.0 (adjusted with NaOH). Store at –20 °C.
2. K_2PdCl_4 stop: 1.5 *M* sodium acetate pH 7.0 and 1 *M* ß-mercaptoethanol.
3. Dimethylsulfate (DMS, 99 + %, Fluka): Considering its toxic and carcinogenic nature, DMS should be manipulated in a well-ventilated hood. DMS is stored under nitrogen at 4 °C and should be replaced every 12 months. DMS waste is detoxified in 5 *M* NaOH.
4. DMS buffer: 50 m*M* sodium cacodylate and 1 m*M* EDTA pH 8. Store at 4 °C.
5. DMS stop: 1.5 *M* sodium acetate pH 7.0 and 1 *M* ß-mercaptoethanol. Store at –20 °C.

6. Hydrazine (Hz, anhydrous, Aldrich): Considering its toxic and carcinogenic potentials, Hz should be manipulated in a well-ventilated hood. Hz is stored under nitrogen at 4 °C in an explosion-proof refrigerator and the bottle should be replaced at least every 6 months. Hz waste is detoxified in 3 *M* ferric chloride.
7. Hz stop: 300 m*M* sodium acetate pH 7.0 and 0.1 m*M* EDTA. Store at 4 °C.
8. 5 *M* NaCl.
9. 3 *M* Sodium acetate pH 7.0.
10. Precooled absolute ethanol (–20 °C).
11. Precooled 80 % ethanol (–20 °C).
12. Dry ice or –80 °C freezer.
13. Piperidine (99 + %, 10 *M*, Fluka or Sigma): Diluted to 2 *M* with H_2O just before use. Cap immediately to minimize evaporation and keep on ice. Considering its toxic and carcinogenic potentials, piperidine should be manipulated in a well-ventilated hood. Piperidine 10 *M* is stored at 4 °C under nitrogen atmosphere.
14. Teflon tape.
15. Lock caps.
16. 3 *M* Sodium acetate pH 5.2.
17. 20 µg/µL glycogen.
18. Vacuum concentrator.

2.3 Template Preparation: PCR Products

2.3.1 PCR Amplification

1. 5× *Taq* buffer: 50 m*M* Tris–HCl pH 8.9, 200 m*M* NaCl, and 0.05 % [w/v] gelatin (*see* **Note 1**).
2. Two primer 2 (50 pmol/µL), one for each strand of the DNA fragment to be amplified, distant from 150 to 450 bp.
3. *Taq* DNA polymerase PCR product mix: 2× *Taq* buffer, 4 m*M* $MgCl_2$, 0.4 m*M* of each dNTP, 10 pmol of each primer 2 and 3 U *Taq* DNA polymerase (5 U/µL, Roche Applied Science).
4. *Taq* DNA polymerase stop: 1.56 *M* sodium acetate pH 5.2 and 60 m*M* EDTA.
5. Precooled absolute ethanol (–20 °C).
6. Precooled 80 % ethanol (–20 °C).
7. Dry ice or –80 °C freezer
8. 5× neutral loading buffer: 0.25 % bromophenol blue, 0.25 % xylene cyanol FF and 30 % glycerol. Store at 4 °C.

2.3.2 Purification and Quantification of PCR Products

1. Agarose.
2. 1× TAE buffer: 40 m*M* Tris base, 20 m*M* glacial acetic acid, and 1 m*M* EDTA pH 8.0.

3. DNA size standard (100 bp, Invitrogen).
4. Ethidium bromide.
5. Glass wool.
6. 3 *M* sodium acetate pH 7.0.
7. Precooled absolute ethanol (–20 °C).
8. Precooled 80 % ethanol (–20 °C).
9. Dry ice or –80 °C freezer.
10. Low DNA mass ladder (Invitrogen).
11. 5× neutral loading buffer: 0.25 % bromophenol blue, 0.25 % xylene cyanol FF, and 30 % glycerol. Store at 4 °C.

2.4 Treatment of Purified DNA and Living Cells with Modifying Agents

2.4.1 DMS Treatment

1. 0.2 % DMS (99+%, Fluka) freshly prepared in serum-free medium.
2. Trypsin-EDTA (Wisent).
3. Hanks' Balanced Salt Solution (HBSS, Wisent), cold.
4. DMS buffer: 50 m*M* sodium cacodylate and 1 m*M* EDTA pH 8. Store at 4 °C.
5. DMS stop: 1.5 *M* sodium acetate pH 7.0 and 1 *M* ß-mercaptoethanol. Store at –20 °C.
6. Buffer A: 300 m*M* sucrose, 60 m*M* KCl, 15 m*M* NaCl, 60 m*M* Tris–HCl pH 8.0, 0.5 m*M* spermidine, 0.15 m*M* spermine, and 2 m*M* EDTA. Store at –20 °C.
7. Buffer A+1 % Nonidet P40 Substitute (Fluka). Store at –20 °C.
8. Buffer B: 150 m*M* NaCl and 5 m*M* EDTA pH 7.8.
9. Buffer C: 20 m*M* Tris–HCl pH 8, 20 m*M* NaCl, 20 m*M* EDTA and 1 % SDS.
10. Proteinase K from *Tritirachium album* (Roche Applied Science).
11. RNase A from bovine pancreas (Roche Applied Science).
12. Phenol, equilibrated, pH 8 (USB Corporation).
13. Chloroform.
14. 5 *M* NaCl.
15. Precooled absolute ethanol (–20 °C).
16. Precooled 80 % ethanol (–20 °C).
17. Dry ice or –80 °C freezer.

2.4.2 UVC (254 nm) and UVB Irradiation

1. Germicidal lamp (254 nm UVC, Philips G15 T8, TUV 15 W).
   or
   UVB light (Philips, FS20T12/UVB/BP).
2. UVX digital radiometer (Ultraviolet Products, Upland, CA).
3. 0.9 % NaCl.

4. 150 mm Petri dishes.
5. UV irradiation buffer: 150 m*M* KCl, 10 m*M* NaCl, 10 m*M* Tris–HCl pH 8.0, and 1 m*M* EDTA.
6. Buffer A: 300 m*M* sucrose, 60 m*M* KCl, 15 m*M* NaCl, 60 m*M* Tris–HCl pH 8.0, 0.5 m*M* spermidine, 0.15 m*M* spermine, and 2 m*M* EDTA. Store at –20 °C.
7. Buffer A + 0.5 % Nonidet P40 Substitute (Fluka). Store at –20 °C.
8. Buffer A + 1 % Nonidet P40 Substitute (Fluka). Store at –20 °C.
9. Scraper.
10. 5 *M* NaCl.
11. Precooled absolute ethanol (–20 °C).
12. Precooled 80 % ethanol (–20 °C).
13. Dry ice or –80 °C freezer.

2.4.3 DNase I Treatment

1. 0.5 mg/mL Deoxyribonuclease I (DNase I, Worthington Biochemical Corporation).
2. Hanks' Balanced Salt Solution (HBSS,Wisent).
3. Solution I: 150 m*M* sucrose, 80 m*M* KCl, 35 m*M* HEPES pH 7.4, 5 m*M* $MgCl_2$, and 0.5 m*M* $CaCl_2$.
4. Solution I + 0.05 %l-a-Lysophosphatidylcholine (l-a-Lysolecithin, Sigma)
5. Solution II: 150 m*M* sucrose, 80 m*M* KCl, 35 m*M* HEPES pH 7.4, 5 m*M* $MgCl_2$, and 2 m*M* $CaCl_2$.
6. Scraper.
7. Conical tubes, 15 mL.
8. Buffer B: 150 m*M* NaCl and 5 m*M* EDTA pH 7.8.
9. Buffer C: 20 m*M* Tris–HCl pH 8.0, 20 m*M* NaCl, 20 m*M* EDTA, and 1 % SDS.
10. Proteinase K from *Tritirachium album* (Roche Applied Science).
11. RNase A from bovine pancreas (Roche Applied Science).
12. Phenol, equilibrated, pH 8 (USB Corporation).
13. Chloroform.
14. 5 *M* NaCl.
15. Glycogen (Roche Applied Science).
16. Precooled absolute ethanol (–20 °C).
17. Precooled 80 % ethanol (–20 °C).
18. Dry ice or –80 °C freezer.

2.5 Conversion of Modified Bases to DNA Single-Strand Breaks

2.5.1 DMS-Induced Base Modifications

1. Piperidine (99 + %, 10 *M*, *see* Subheading 2.2, **item 13**).
2. Teflon tape.
3. Lock caps.
4. 3 *M* Sodium acetate pH 5.2.
5. 20 μg/μL glycogen (Roche Applied Science).
6. Precooled absolute ethanol (−20 °C).
7. Precooled 80 % ethanol (−20 °C).
8. Dry ice of −80° freezer.
9. Vacuum concentrator.

2.5.2 UV-Induced Base Modifications

1. 10× dual buffer: 500 m*M* Tris–HCl pH 7.6, 500 m*M* NaCl, and 10 m*M* EDTA.

CPD

2. T_4 endonuclease V enzyme (Trevigen): The saturating amount of T_4 endonuclease V enzyme can be estimated by digesting UV-irradiated genomic DNA with various enzyme quantities and separating the cleavage products on alkaline agarose gel [47]. The saturating amount of the enzyme is the one next to the minimum quantity that produces the maximum cleavage frequency as evaluated on the alkaline agarose gel.
3. T_4 endo V mix: 2× dual buffer, 2 m*M* DTT (1,4-Dithiothreitol, Roche Applied Science), 0.2 mg/mL BSA (nuclease-free bovine serum albumin, Roche Applied Science) and a saturating amount of T_4 endonuclease V.
4. *E. coli* photolyase enzyme (Trevigen). The saturating amount of photolyase can be estimated by photoreactivating UV-irradiated genomic DNA with various enzyme quantities, digestion with T_4 endonuclease V and separating the cleavage products on alkaline agarose gel [47]. The saturating amount of photolyase is the next to the minimum enzyme quantity which produces no cleavage following T_4 endonuclease V digestion as evaluated on the gel. Because photolyase is light sensitive, all steps involving photolyase should be carried out rapidly.
5. Photolyase mix: 1× dual buffer, 1.1 *M* DTT, 0.1 mg/mL BSA, and a saturating amount of photolyase.
6. UVA black light blue (Sankyo Denki 350 nm).
7. Plastic film (plastic wrap).
8. 1 % SDS solution.
9. Phenol, equilibrated, pH 8 (USB Corporation).
10. Chloroform.
11. 5 *M* NaCl.
12. Precooled absolute ethanol (−20 °C).
13. Precooled 80 % ethanol (−20 °C).
14. Dry ice of −80 °C freezer.

6-4PP

1. Piperidine (99 + %, 10 *M*, *see* Subheading 2.2, **item 13**).
2. Teflon tape.
3. Lock caps.
4. 3 *M* Sodium acetate pH 5.2.
5. 20 µg/µL glycogen (Roche Applied Science).
6. Precooled absolute ethanol (–20 °C).
7. Precooled 80 % ethanol (–20 °C).
8. Dry ice of –80° freezer.
9. Vacuum concentrator.

2.6 Ligation-Mediated Polymerase Chain Reaction Technology

2.6.1 Primer Extension (Steps II and III, Fig. 5)

1. A gene-specific primer (primer 1) is used to initiate PE. The primer 1 used in the first-strand synthesis is 15–22 bp oligonucleotides and have a calculated melting temperature (T_m) of 50–60 °C. They are selected using a computer program (Oligo 4.0 software, National Biosciences, Rychlik and Rhoads 1989) and optimally, their T_m, as calculated by a computer program (GeneJockey software), should be about 10 °C lower than that of subsequent primers (*see* **Note 2**) [48]. The first-strand synthesis reaction is designed to require very little primer 1 with a lower T_m so that this primer does not interfere with subsequent steps [11–13, 49]. The primer 1 concentration is set at 0.5 pmol/µL in H_2O.
2. Thermocycler (Biometra or PTC™, MJ research, Inc.).
3. Polymerase extension mix:

3a. If *Pfu* exo- DNA polymerase (also named cloned *Pfu*) is used:

- 10× *Pfu* exo- buffer: 200 m*M* Tris–HCl pH 8.8, 20 m*M* $MgSO_4$, 100 m*M* NaCl, 100 m*M* $(NH_4)_2SO_4$, 1 % (v/v) Triton X-100, and 1 mg/mL nuclease-free BSA (*see* **Note 1**).
- *Pfu* exo- extension mix: 1.25 pmole primer 1, 0.25 m*M* of each dNTP, 1× *Pfu* exo- buffer, and 1.5 U *Pfu* exo$^-$ (2.5 U/µL, Stratagene).

3b. If *Vent* exo- polymerase is used:

- *Vent* exo- extension mix: 1.25 pmole primer 1, 0.25 m*M* of each dNTP, 1× *Vent* buffer (10×, New England BioLabs), and 0.75 U *Vent* exo- polymerase (2.0 U/µL New England BioLabs).

2.6.2 Ligation (Step IV, Fig. 5)

1. The DNA molecules that have a 5′-phosphate group and a double-stranded blunt end are suitable for ligation. A DNA linker with a single blunt end is ligated directionally onto the double-stranded blunt end of the extension product using T_4 DNA ligase. This linker has no 5′ phosphate and is staggered to avoid self-ligation and provide directionality. Also, the duplex between the 25-mer (L25: 5′ GCGGTGACCCGGG AGATCTGAATTC) and 11-mer (L11: 5′ GAATTCAGATC)

is stable at the ligation temperature, but denatures easily during subsequent PCR reactions [6, 49]. The linker 20 pmol/µL is prepared in aliquots of 500 µL by annealing in 250 m*M* Tris–HCl pH 7.7, 120 m*M* $MgCl_2$, and 20 pmol/µL each of the 25-mer and 11-mer (the stock is at 60 pm/µL), heating at 95 °C for 3 min, transferring quickly at 70 °C, cooling gradually to room temperature, and storing at 4 °C overnight. Linker is stored at –20 °C and thawed on ice before use.

2. Ligation mix: 33 m*M* DTT, 1.1 m*M* ATP, 16.6 µg/mL BSA, 48.9 m*M* Tris–HCl pH 7.4, 100 pmol linker, and 3.25 U T_4 ligase (1 U/µL, Roche Applied Science).
3. T_4 ligase stop mix: 7.2 *M* Ammonium acetate, 4.2 m*M* EDTA pH 8.0, and 0.67 µg/µL glycogene (Roche Applied Science).
4. Precooled absolute ethanol (–20 °C).
5. Precooled 80 % ethanol (–20 °C).
6. Dry ice or –80 °C freezer.

2.6.3 Polymerase Chain Reaction (Steps V and VI, Fig. 5)

1. At this step, gene-specific fragments can be exponentially amplified because primer sites are available at each target fragment ends (i.e., primer 2 on one end and the longer oligonucleotide of the linker on the other end). Primer 2 may or may not overlap with primer 1. The overlap, if present, should not be more than seven to eight bases [11–13, 49]. The primer 2 and the linker primer (L25) are diluted in H_2O to give 50 pmol/µL and 60 pmol/µL, respectively.
2. Thermocycler (Biometra or PTC™, MJ research, Inc.).
3. Polymerase amplification reaction.

3a. If *Pfu* exo$^-$ polymerase is used:

- 10× *Pfu* exo- buffer: 200 m*M* Tris–HCl pH 8.8, 20 m*M* $MgSO_4$, 100 m*M* NaCl, 100 m*M* $(NH_4)_2SO_4$, 1 % (v/v) Triton X-100, and 1 mg/mL nuclease-free BSA (*see* **Note 1**).
- *Pfu* exo- amplification mix: 2× *Pfu* exo- buffer, 0.5 m*M* of each dNTP, 10 pmol of L25 (linker primer), 10 pmol of primer 2, and 3.5 U of *Pfu* exo- DNA polymerase (2.5U/µL, Stratagene).

3b. If *Taq* DNA polymerase is used:

- 5× *Taq* buffer: 50 m*M* Tris–HCl pH 8.9, 200 m*M* NaCl, and 0.05 % [w/v] gelatin (*see* **Note 1**).
- *Taq* amplification mix: 2× *Taq* buffer, 4 m*M* $MgCl_2$, 0.5 m*M* of each dNTP, 10 pmol L25 (linker primer), 10 pmol primer 2, and 3 U *Taq* DNA polymerase (5U/µL, Roche Applied Science).

3c. If *Vent* exo- is used:

– *Vent* exo- amplification mix: 2× *Vent* exo- buffer (10×, New England BioLabs), 0.5 m*M* of each dNTP, 10 pmol of L25 (linker primer), 10 pmol of primer 2, and 2.5 U *Vent* exo- polymerase (2.0 U/μL New England BioLabs).

2.7 LMPCR-Amplified DNA Fragment Analysis

2.7.1 Conventional Radioactive Method

Precipitation, Gel Electrophoresis, and Electroblotting (Step VII, Fig. 5)

1. Stop mix:

1a. If *Pfu* exo- is used for amplification:

– *Pfu* exo- stop mix: 1.56 *M* sodium acetate pH 5.2 and 20 m*M* EDTA.

1b. If *Taq* DNA polymerase is used for amplification:

– *Taq* stop mix: 1.56 *M* sodium acetate pH 5.2 and 60 m*M* EDTA.

1c. If *Vent* exo- is used for amplification:

– *Vent* exo- stop mix: 1.56 *M* sodium acetate pH 5.2 and 20 m*M* EDTA.

2. Precooled absolute ethanol (–20 °C).
3. Precooled 80 % ethanol (–20 °C).
4. Dry ice or –80 °C freezer.
5. Formamide loading dye: 94 % formamide, 2 m*M* EDTA pH 7.7, 0.05 % xylene cyanol FF, and 0.05 % bromophenol blue [11–13].
6. 60-cm-long × 34.5-cm-wide sequencing gel apparatus (Owl Scientific).
7. Spacers (0.4 mm thick).
8. Plastic well-forming comb (0.4 mm thick, BioRad).
9. 5× (0.5 *M*) Tris-borate-EDTA (TBE) buffer: 500 m*M* Tris, 830 m*M* boric acid and 10 m*M* EDTA pH 8.3. Use this stock to prepare 1× (100 m*M*) TBE buffer.
10. 8 % polyacrylamide. To prepare 1 L: 77.3 g acrylamide, 2.7 g *bis*-acrylamide, 420.42 g urea, and 200 mL of 0.5 *M* TBE dissolved in H_2O. Polyacrylamide solution should be kept at 4 °C.
11. Gel preparation: Mix 100 mL of 8 % polyacrylamide with 1 mL of 10 % ammonium persulfate (APS) and 30 μL of *N,N,N′,N′*-tetra-methylethylenediamide (TEMED). This mix is prepared immediately before pouring the solution between the glass plates. Without delay, take the gel mix into a 50 mL syringe and inject the mix between the plates, maintaining a steady flow. During pouring, the plates should be kept at a 30° angle and tilted to the side into which the mix is injected. Any air bubbles should be avoided and removed if they form. The gel should be left to polymerize for a minimum of 2 h before use. If the gel is to be left overnight, 45 min after pouring, place a

moistened paper tissue over the comb and cover the upper end of the assembly with a plastic film to prevent the gel from drying out.

12. Power supply (Bio-Rad PowerPac 3000).
13. Electroblotting apparatus (HEP3, Owl Scientific Inc.) used according to the manufacturer's instructions.
14. Whatman 3 MM Chr paper (Fisher Scientific).
15. Plastic film (plastic wrap).
16. Whatman 17 MM Chr paper (Fisher Scientific).
17. Nylon membrane, positively charged (Roche Applied Science).
18. Power supply (Bio-Rad, model 200/2.0).
19. UVC (254 nm) germicidal lamp.
20. UVX digital radiometer (Ultraviolet Products, Upland, CA).

Preparation of Single-Stranded Hybridization Probes (Step VIII, Fig. 5)

1. 5× *Taq* buffer: 50 m*M* Tris–HCl pH 8.9, 200 m*M* NaCl, and 0.05 % [w/v] gelatin (*see* **Note 1**).
2. dNTP (dATP, dGTP, dTTP) mix (200 µ*M* of each) diluted 1:10 in H_2O. This mix is changed every 2 weeks.
3. Isotopic labeling mix: 1× *Taq* buffer, 2 m*M* $MgCl_2$, 133 pg/µL DNA template (PRC products), 0.5 pm/µL primer 2, 1 µL of 1:10 dNTP mix, 5U *Taq* DNA polymerase (5 U/µL, Roche Applied Science), and 1.85 MBq α-[^{32}P] dCTP (0.37 MBq/µL, PerkinElmer Life Sciences Inc).
4. 7.5 *M* Ammonium acetate.
5. 20 µg/µL glycogen.
6. Precooled absolute ethanol (–20 °C).
7. Geiger counter.
8. TE buffer pH 8.0: 10 mM Tris–HCl pH 8.0 and 1 mM EDTA pH 7.8.
9. Hybridization buffer: 250 m*M* sodium phosphate pH 7.2, 1 m*M* EDTA, 7 % SDS, and 1 % BSA.

Hybridization (Step VII, Fig. 5), Washing, and Autoradiography

The hybridization is performed in a rolling 8 cm diameter × 22 cm long borosilicate glass hybridization tubes in a hybridization oven (Techne). The nylon membrane is soaked in 100 m*M* TBE and placed in the tube using a 25 mL pipet, so that the membrane sticks completely to the wall of the hybridization tube. Following hybridization and washing, the membrane is placed in an autoradiography cassette Fujifilm EC-DW (Christie Group Ltd) and exposed to Kodak X-ray film (BiomaxMR, 35 × 43 cm, Kodak Scientific Imaging Film) with intensifying screens (35 × 43 cm, Fisher Scientific, cat. no. FB-IS-1417) at –80 °C.

1. Hybridization buffer: 250 m*M* sodium phosphate pH 7.2, 1 m*M* EDTA, 7 % SDS, and 1 % BSA.
2. Single-stranded hybridization probe diluted in 6–7 mL of hybridization buffer.
3. Washing buffer I: 20 m*M* sodium phosphate pH 7.2, 1 m*M* EDTA, 0.25 % BSA, and 2.5 % SDS.
4. Washing buffer II: 20 m*M* sodium phosphate pH 7.2, 1 m*M* EDTA, and 1 % SDS.
5. Plastic film (plastic wrap).
6. Kodak X-ray film (BiomaxMR, 35×43 cm, Kodak Scientific Imaging Film).
7. Autoradiography cassette Fujifilm EC-DW (Christie Group Ltd).
8. Intensifying screens (35×43 cm, Fisher Scientific).

2.7.2 Fluorescent Method Using Sequencer

Fluorescent Labeling

1. At this step, in order to remove excess of free primer 2, an exonuclease digestion is performed, followed by a fluorescent labeling extension of all DNA fragments amplified during the LMPCR step. For this purpose a third primer (primer 3) fluorescently labeled is used to perform five consecutive extension cycles. This primer is located right after the 3′ end of the primer 2. The primer 3 is diluted in H_2O to give 1 nmol/mL.
2. Thermocycler (Biometra or PTC™, MJ research, Inc.).
3. Exonuclease I (20,000 U/mL, New England Biolabs).
4. 5× *Taq* buffer: 50 m*M* Tris–HCl pH 8.9, 200 m*M* NaCl, and 0.05 % [w/v] gelatin (*see* **Note 1**).
5. Exonuclease I mix: 1× *Taq* buffer, 667 U/mL Exonuclease I.
6. 1 μ*M* primer 3, fluorescent labeled (Li-Cor Bioscience).
7. Labeling stop mix: 1.56 *M* sodium acetate pH 5.2 and 60 mM EDTA.
8. Precooled absolute ethanol (–20 °C).
9. Precooled 80 % ethanol (–20 °C).
10. Dry ice or –80 °C freezer.
11. Formamide loading dye: 94 % formamide, 2 m*M* EDTA pH 7.7, 0.05 % xylene cyanol FF, and 0.05 % bromophenol blue [11–13].

Sequencing Gel

1. 66-cm-long sequencing gel apparatus (LI-COR Bioscience).
2. Spacers (0.2 mm thick, LI-COR Bioscience).
3. Plastic well-forming comb (0.2 mm thick, LI-COR Bioscience).
4. 5× (0.5 *M*) TBE buffer: 500 m*M* Tris, 830 m*M* boric acid, and 10 m*M* EDTA pH 8.3. Use this stock to prepare 1× (100 m*M*) TBE buffer.

5. 8 % polyacrylamide. To prepare 1 L: 77.3 g acrylamide, 2.7 g *bis*-acrylamide, 420.42 g urea, and 200 mL of 0.5 *M* TBE dissolved in H_2O. Polyacrylamide solution should be kept at 4 °C.
6. Gel preparation: Mix 50 mL of 8 % polyacrylamide with 0.5 mL of 10 % APS and 15 μL of TEMED. This mix is prepared immediately before pouring the solution between the glass plates. Without delay, take the gel mix into a 50 mL syringe and inject the mix between the plates, maintaining a steady flow. During pouring, the plates should be kept at a 30° angle. Any air bubbles should be avoided and removed if they form. The gel should be left to polymerize for a minimum of 2 h before use. If the gel is to be left overnight, 45 min after pouring, place a moistened paper tissue over the comb and cover the upper end of the assembly with a plastic film to prevent the gel from drying out.
7. DNA 4300 sequencer (LI-COR Bioscience).

3 Methods

3.1 DNA Purification (for 10^7 to 10^8 Cells)

1. Detach cells using trypsin (if needed) and sediment the cell suspension by centrifugation in 50 mL conical tubes.
2. Resuspend the cells in 2–8 mL of buffer A.
3. Add 1 volume (2–8 mL) of buffer A containing 1 % Nonidet P40 Substitute.
4. Incubate at 4 °C for 5 min.
5. Sediment nuclei by centrifugation at 4500 × *g* for 15 min at 4 °C.
6. Remove the supernatant. Resuspend nuclei in 5–10 mL of buffer A by gentle vortexing. Resediment nuclei at 4500 × *g* for 15 min at 4 °C.
7. Remove supernatant. It is recommended to leave a small volume (100–500 μL) of buffer A to facilitate resuspension of nuclei.
8. Dilute the nuclei in 1–2 mL of buffer B.
9. Add an equivalent volume of buffer C and proteinase K to a final concentration of 450 μg/mL.
10. Incubate at 37 °C for 3 h (*see* **Note 3**).
11. Add RNase A to a final concentration of 150 μg/mL.
12. Incubate at 37 °C for 1 h.
13. Purify DNA by extraction with 1 volume phenol (one or two times as needed), 1 volume phenol:chloroform (one or two times as needed), and 1 volume chloroform. Phenol extraction

and phenol-chloroform extraction should be repeated if the aqueous phase is not clear .

14. Precipitate DNA in 200 m*M* NaCl and 2 volumes of precooled absolute ethanol. Ethanol should be added slowly. Make sure to mix very gently.
15. Recover DNA by spooling the floating DNA filament with a micropipet tip. If DNA is in small pieces or not clearly visible, recover DNA by centrifugation (4500 × *g* for 15 min at 4 °C), but expect RNA contamination (*see* **Note 4**). RNA contamination does not cause any problems for LMPCR. RNase digestion can be repeated if needed.
16. Wash DNA once with 10 mL of 80 % ethanol.
17. Centrifuge the DNA (4500 × *g* for 10 min at 4 °C).
18. Remove supernatant and air-dry DNA pellet.
19. Dissolve DNA in water at an estimated concentration of 60–100 µg/mL. The quantity of DNA can be estimated based upon the number of cells that were initially used for DNA purification. About 6 µg of DNA should be purified from 1×10^6 cells.
20. Carefully measure DNA concentration by spectrophotometry at 260 nm. Alternatively, DNA can be measured by fluorometry after staining with DAPI. Only double-stranded DNA concentration must be measured; be careful if there is RNA contamination (*see* **Note 5**).

3.2 Chemical Cleavage for DNA Sequencing Products

In cellulo DNA analysis using LMPCR requires complete DNA sequencing ladders from genomic DNA. Base-specific chemical modifications are performed according to Iverson and Dervan [50] for the A reaction and Maxam and Gilbert for the G, T + C and C reactions. DNA from each of these base modification reactions is processed by LMPCR concomitantly with the analyzed samples and loaded in adjacent lanes on the sequencing gel to allow the identification of the precise location and sequence context of footprinted regions. The chemical modifications induced by DMS, Hz, and K_2PdCl_4 and cleaved by piperidine destroy the target base. Therefore, one must bear in mind that when analyzing a chemical-sequencing ladder, each band corresponds to a DNA fragment ending at the base preceding the one read. In this section, we describe the chemical sequencing of genomic DNA. The cleavage protocol below works optimally with 10–50 µg of genomic DNA per microtube. The required amount of DNA is ethanol precipitated and the pellet is air-dried. For each base-specific reaction, we usually carried out the treatment in three microtubes containing 50 µg of genomic DNA for three different incubation times with the modifying agent in order to obtain low, medium, and high base-modification frequencies.

3.2.1 A Reaction

1. Add 160 µL of H_2O to dissolve the DNA pellet.
2. Add 40 µL of K_2PdCl_4 solution, mix carefully, and keep on ice.
3. Incubate at room temperature for 5, 10, or 15 min.
4. Add 50 µL of K_2PdCl_4 stop.
5. Add 750 µL of precooled absolute ethanol.

3.2.2 G Reaction

1. Add 5 µL of H_2O to dissolve the DNA pellet and mix.
2. Add 200 µL of DMS buffer and 1 µL of DMS, carefully mix, and keep on ice.
3. Incubate at room temperature for 30, 45, or 60 s.
4. Add 50 µL of DMS stop.
5. Add 750 µL of precooled absolute ethanol.

3.2.3 T+C Reaction

1. Add 20 µL of H_2O to dissolve the DNA pellet and mix.
2. Add 30 µL of Hz, carefully mix, and keep on ice.
3. Incubate at room temperature for 120, 210, or 300 s.
4. Add 200 µL of Hz stop.
5. Add 750 µL of precooled absolute ethanol.

3.2.4 C Reaction

1. Add 5 µL of H_2O to dissolve the DNA pellet and mix .
2. Add 15 µL of 5 *M* NaCl and 30 µL of Hz, carefully mix, and keep on ice.
3. Incubate at room temperature for 120, 210, or 300 s.
4. Add 200 µL of Hz stop.
5. Add 750 µL of precooled absolute ethanol.

All samples are processed as follows:

6. Mix samples well and place on dry ice for 15 min or at –80 °C until the samples are frozen.
7. Centrifuge for 15 min at 16,000 × *g*.
8. Remove supernatant, recentrifuge for 1 min, and remove all the liquid using a micropipet.
9. Carefully dissolve pellet in 405 µL of H_2O.
10. Add 45 µL of 3 *M* sodium acetate pH 7.0.
11. Add 1 mL of precooled absolute ethanol and mix well.
12. Leave on dry ice for 15 min or at –80 °C until the samples are frozen.
13. Centrifuge for 15 min at 16,000 × *g*.
14. Remove supernatant.
15. Wash with 1 mL of precooled 80 % ethanol and centrifuge for 5 min at 16,000 × *g*.

16. Remove the supernatant, spin quickly, remove the liquid with a micropipet, and air-dry pellet.
17. Dissolve pellet in 50 µL H_2O. Add 50 µL of freshly prepared 2 *M* piperidine and mix well.
18. Secure caps with Teflon tapes and lock the caps with "lock caps."
19. Incubate at 80 °C for 30 min.
20. Pool all three microtubes of the same chemical reaction in a new 1.5 mL microtube.
21. Add 105 µL H_2O, 45 µL of 3 *M* sodium acetate pH 5.2, 1 µL of glycogen, and 1 mL of precooled absolute ethanol and mix well.
22. Leave on dry ice for 15 min or at –80 °C until the samples are frozen.
23. Spin 15 min at 16,000 × *g*.
24. Take out the supernatant and wash twice with 1 mL of precooled 80 % ethanol, respin for 5 min, and remove all the liquid using a micropipet.
25. Add 200 µL of H_2O and remove traces of remaining piperidine by drying the sample in a Speedvac concentrator.
26. Dissolve DNA in H_2O to a concentration of 0.5 µg/µL.
27. Determine the DNA strand break frequency by running the samples on a 1.5 % alkaline agarose gel [47]. The size range of the fragments should span 100–500 bp.

3.3 Template Preparation: PCR Products

3.3.1 PCR Amplification

1. To 100 ng of purified genomic DNA in H_2O, add 50 µL of the *Taq* DNA polymerase PCR product mix and mix.
2. Cycle 35 times at 95 °C for 1 min (97 °C for 3 min for the first cycle), 61–73 °C (1–2 °C below the calculated T_m of the primer 2 with the lowest T_m) for 2 min, and 74 °C for 3 min. The last extension should be done for 10 min.
3. Add 25 µL of *Taq* DNA polymerase stop.
4. Add 400 µL of precooled absolute ethanol and mix well.
5. Leave for 15 min on dry ice or at –80 °C until the samples are frozen, and spin for 5 min at 16,000 × *g*.
6. Wash once with 1 mL of precooled 80 % ethanol.
7. Spin for 5 min at 16,000 × *g*.
8. Air-dry DNA pellets.
9. Resuspend DNA pellets in 12 µL H_2O and add 3 µL of 5× neutral loading buffer.

3.3.2 Purification and Quantification of PCR Products

1. Load 15 µL of PCR products per well along with a 100 bp DNA size standard.
2. Migrate the PCR products on a 1 % neutral agarose gel.
3. Stain the gel with ethidium bromide and photograph on a UV transilluminator. Recover the band containing the DNA fragment of expected molecular weight using a clean scalpel blade. Minimize the size of the slice by removing as much extraneous agarose as possible.
4. Crush the slice and put it in a 0.6 mL microtube pierced at the bottom, and containing a column of packed dry glass wool (*see* **Note 6**).
5. Insert the 0.6 mL microtube containing the column in a 1.5 mL microtube and spin for 15 min at 7000 × *g*. Transfer the flow-through to a new 1.5 mL microtube. If there is still some agarose remaining, respin for 15 min at 7000 × *g*.
6. Add 50 µL of H_2O to wash the column of any remaining DNA by spinning for 8 min at 7000 × *g*. Pool all of the flow-through contents in one 1.5 mL microtube.
7. Complete the volume to 405 µL with H_2O, add 45 µL of 3 *M* sodium acetate pH 7.0 and 1 mL of precooled absolute ethanol to precipitate DNA. Mix well and leave for 15 min on dry ice or at −80 °C until the samples are frozen. Spin for 15 min at 16,000 × *g*.
8. Wash once with 1 mL of precooled 80 % ethanol and spin for 5 min at 16,000 × *g*.
9. Air-dry DNA pellet.
10. Dissolve DNA pellets in 104 µL H_2O.
11. On a 1.5 % neutral agarose gel, load aliquots of 1 and 3 µL of the DNA template dissolved in 1× neutral loading buffer along with a quantitative low DNA mass ladder.
12. Stain the gel with ethidium bromide and photograph on a UV transilluminator. The DNA concentration of the aliquots is estimated by comparison with the low DNA mass ladder band intensities and H_2O is added to obtain a final concentration of template DNA of 10 ng/µL. The DNA template is aliquoted and stored at −20 °C.

3.4 Treatment of Purified DNA and Cells with Modifying Agents

3.4.1 DMS Treatment

1. If cells are grown to confluence as a monolayer, replace the culture medium with a freshly prepared serum-free medium containing 0.2 % DMS and incubate at room temperature for 6 min. If cells are grown in suspension, sediment the cells by centrifugation and remove the cell culture medium. The cells are diluted in a freshly prepared serum-free medium containing 0.2 % DMS and are then incubated at room temperature for 6 min.

2. Remove the DMS-containing medium and quickly wash the cell monolayer with 10 mL of cold HBSS. Sediment cells by centrifugation if they are treated in suspension and remove the DMS-containing medium and wash the cells with 10 mL of cold HBSS.
3. Detach cells using trypsin for cells grown as monolayer.
4. Nuclei are isolated and DNA purified as described in Subheading 3.1.
5. Purified DNA obtained from the same cell type is treated as described in Subheading 3.2.2. Usually, a DMS treatment of 45 s should give a break frequency corresponding to that of the in cellulo treatment described in this section. This DNA is the in vitro-treated DNA used to compare with DNA DMS-modified in cellulo (*see* **Notes 5** and 7).

3.4.2 UVC (254 nm) and UVB Irradiation

1. If cells are grown as monolayer in Petri dishes, replace cell culture medium with cold 0.9 % NaCl. If cells are grown in suspension, sediment the cells by centrifugation and remove the cell culture medium. The cells are diluted in cold 0.9 % NaCl at a concentration of 1×10^6 cells/mL (*see* **Note 8**) and, to avoid cellular shielding, a thin layer of the cell suspension is placed in 150 mm Petri dishes.
2. Expose the cells to 0.5–2 kJ/m^2 of UVC (254 nm UV) or 25–100 kJ/m^2 of UVB. The cells should be exposed on ice with uncovered Petri dishes. The UV intensity is measured using a UVX digital radiometer.
3. Remove the 0.9 % NaCl by aspiration for cells grown as monolayer in Petri dishes or by sedimentation for cell suspensions.
4. If cells were irradiated in suspension, follow the procedure described in Subheading 3.1 to isolate nuclei and purify DNA. After DNA purification, DNA is dissolved in H_2O at a concentration of 0.2 μg/μL. For cells cultured in Petri dishes, add in each dish 8 mL of buffer A containing 0.5 % Nonidet P40 Substitute.
5. Incubate at 4 °C for 5 min.
6. Scrape the cells and transfer them in a conical 50 mL tube.
7. Wash the dishes once with 8 mL of buffer A + 0.5 % Nonidet P40 Substitute.
8. Continue from **step 5** of Subheading 3.1. After DNA purification, DNA is dissolved in H_2O at a concentration of 0.2 μg/μL.
9. Expose purified DNA to the same UVC or UVB dose as the cells. Purified DNA should be irradiated on ice and diluted in the UV irradiation buffer at a concentration of 60–75 μg/mL (*see* **Note 7**). Purified DNA should be obtained from the same

type of cells as the type irradiated in cellulo (*see* **Note 9**). This DNA is used as control DNA to compare with DNA UV modified in cellulo (*see* **Notes 7** and **8**).

10. Following UV irradiation, DNA is ethanol precipitated and DNA is resuspended in H_2O at a concentration of 0.2 μg/μL.

3.4.3 DNase I Treatment

Genomic footprinting with DNase I requires cell permeabilization (*see* **Note 10**). Cells grown as a monolayer can be permeabilized while they are still attached to the Petri dish or in suspension following trypsinization. Here, we will describe cell permeabilization using lysolecithin applied to monolayer cell cultures (alternatively, the cells can be permeabilized using Nonidet P40 Substitute, see detailed protocol in [18]). For monolayer cultures, cells are grown to about 80 % of confluency. For cells in suspension, cells are diluted at a concentration of approximately 1×10^6 cells/mL. To permeabilize the vast majority of cells in suspension, they must not be clumped and must not form aggregates during the permeabilization step and subsequent DNase I treatment. To achieve this, we gently flick the microtubes during permeabilization and DNase I treatment and keep the cell concentration below 2×10^6/mL.

1. For cells in monolayers, wash the cells with 5 mL solution I and permeabilize the cells by treating them with 4 mL of solution I + 0.05 % lysolecithin at 37 °C [51].
2. Add 55 μL of DNase I (0.5 mg/mL). Incubate at 37 °C for 20 min. DNase I concentration and incubation times may have to be adjusted for different cell types.
3. After 8 min, detach cells using scraper and transfer in a 15 mL tubes. Replace at 37 °C for the rest of the 20-min incubation.
4. Centrifuge for 1 min at 450 × *g*.
5. Resuspend the pellet in 500 μL buffer B.
6. Add 500 μL buffer C.
7. Add 450 μg/mL Proteinase K and incubate at 37 °C for 3 h.
8. Add RNase A to a final concentration of 200 μg/mL and incubate at 37 °C for 1 h.
9. Purify DNA by phenol-chloroform extraction (see Subheading 3.1, **step 13**).
10. Precipitate DNA in 200 m*M* NaCl, 1 μL glycogen and 2 volumes of precooled absolute ethanol.
11. Recover DNA by centrifugation (4500 × *g* for 15 min at 4 °C), but expect RNA contamination. RNA contamination does not cause any problems for LMPCR. RNase A digestion can be repeated if needed.

12. Remove supernatant and wash DNA once with 10 mL of precooled 80 % ethanol.
13. Centrifuge the DNA (4500 × *g* for 10 min at 4 °C). Remove supernatant and air-dry DNA pellet.
14. Dissolve DNA in H_2O and measure DNA concentration (*see* Subheading 3.1, **step 20**).
15. To obtain purified DNA controls (*see* **Notes 7** and **9**), digest 40 μg of purified DNA in solution II with 10–20 ng/mL of DNase I at room temperature for 10–20 min. Stop the reaction by adding 400 μL of phenol. Extract once with phenol-chloroform and once with chloroform. Dissolve DNA in H_2O at a concentration of 0.2 μg/μL.

3.5 Conversion of Modified Bases to DNA Single-Strand Breaks

When purified DNA or living cells are treated with DMS or UV, DNA base modifications are induced (Table 3). These modifications must be converted to single-strand breaks before running LMPCR. Following UV exposure, CPD and 6-4PP are converted individually because they use different conversion procedures (Table 3). On the other hand, DNase I digestion directly generates DNA strand breaks suitable for LMPCR without any conversion procedures. Before running LMPCR, the DNA strand break frequency must be determined by running the samples on a 1.5 % alkaline agarose gel [48]. The size range of the fragments should span 200–2000 bp (*see* **Note 7**).

3.5.1 DMS-Induced Base Modifications (See *Fig. 1*)

1. Dissolve DNA (10–50 μg) in 50 μL H_2O, add 50 μL of 2 *M* piperidine and mix well.
2. Samples are processed as described in Subheading 3.2, **steps 18–27**.
3. Dissolve DNA in H_2O to a concentration of 0.2 μg/μL.

3.5.2 UV-Induced Base Modifications

CPD (*See* Fig. 2)

1. To specifically cleave CPD, dissolve 10 μg of UV-irradiated DNA in 50 μL H_2O and add 50 μL of T_4 endo V mix.
2. Incubate at 37 °C for 1 h.
3. Perform the photolyase digestion to remove the overhanging dimerized base that would otherwise prevent ligation [8], and add 10 μL of the photolyase mix.
4. Leaving their caps opened, cover the microtubes with a plastic film to prevent UVB-induced damage and place open ends 2–3 cm from a UVA black light blue for 1 h.
5. Add 200 μL of 1 % SDS and 100 μL H_2O, mix well.
6. Extract DNA using 1 volume (400 μL) phenol, 1 volume phenol:chloroform, and 1 volume chloroform.
7. To precipitate DNA, add 20 μL of 5 *M* NaCl and 1 mL of precooled absolute ethanol and mix well.

8. Leave for 15 min on dry ice or at –80 °C until the samples are frozen, and spin for 15 min at 16,000 × *g*.
9. Wash once with 1 mL of precooled 80 % ethanol.
10. Spin for 5 min at 16,000 × *g*.
11. Air-dry the pellet and dissolve DNA in H_2O to a concentration of 0.2 μg/μL.

6-4PP (*See* Fig. 3)

1. Dissolve DNA (10–50 μg) in 50 μL of H_2O, add 50 μL of 2 *M* piperidine, and mix well.
2. Samples are processed as described in Subheading 3.2, **steps 18–27**.
3. Dissolve DNA in H_2O to a concentration of 0.2 μg/μL.

3.6 Ligation-Mediated Polymerase Chain Reaction Technology

The PE, ligation, and PCR steps are carried out in 0.6 mL microtubes and a thermocycler is used for all incubations.

3.6.1 Primer Extension (Steps II and III, Fig. 5) (See Note 11)

1. Mix 0.5–2 μg of genomic DNA with the appropriate polymerase extension mix (*Pfu* exo- or *Vent* exo- extension mix) for a final volume of 30 μL.
2. Denature DNA at 98 °C. Incubate the samples at the annealing temperature for 4 min, and then incubate at 75 °C for 10 min. Finally, the samples are cooled to 4 °C.

3.6.2 Ligation (Step IV, Fig. 5)

1. To the primer extension reaction, add 45 μL of the ligation mix and mix well.
2. Incubate at 18 °C for a minimum of 2 h.
3. Precipitate DNA by adding 30 μl of T_4 ligase stop mix and 300 μL of precooled absolute ethanol and mix well.
4. Leave for 15 min on dry ice or at –80 °C until the samples are frozen and spin for 15 min at 16,000 × *g*.
5. Wash once with 500 μL of precooled 80 % ethanol.
6. Spin for 5 min at 16,000 × *g*.
7. Remove supernatant and air-dry DNA pellets.
8. Dissolve DNA in 50 μL H_2O.

3.6.3 Polymerase Chain Reaction (Steps V and VI, Fig. 5) (See Note 11*)*

1. Add 50 μL of the appropriate polymerase amplification mix (*Pfu* exo-, *Vent* exo- or *Taq* amplification mix) and mix.
2. Cycle 22 times as described in Table 4. The last extension should be done for 10 min to fully extend all DNA fragments.

Table 4
Exponential amplification steps using cloned *Pfu* DNA polymerase or *Taq* DNA polymerase

Cycles	Denaturation (T in °C for D in s)		Annealing (T is the T_m of the oligonucleotide for D in s)	Polymerization (D in s) T is the same for all cycles: 75 °C for *Pfu* and 74 °C for *Taq*
	Pfu	*Taq*	*Pfu* or *Taq*	–
0	–	93 for 120	–	–
1	98 for 300	98 for 150	T_m for 180	180
2	98 for 120	95 for 60	T_m for 150	180
3	98 for 60	95 for 60	T_m –2 °C for 120	180
4	98 for 30	95 for 60	T_m –3 °C for 120	180
5	98 for 20	95 for 60	T_m –4 °C for 90	150
Repeat cycle 5, 13 more times (add 5 s per cycle for annealing and polymerization)				
6	98 for 20	95 for 60	T_m –3 °C for 240	240
7	98 for 20	95 for 60	T_m –2 °C for 240	240
8	98 for 20	95 for 60	T_m –1 °C for 240	240
9	98 for 20	95 for 60	T_m for 120	600

Note: Temperature (*T*) and duration (D) of the denaturation, annealing and polymerization steps

3.7 LMPCR-Amplified DNA Fragment Analysis

3.7.1 Conventional Radioactive Method

Precipitation, Gel Electrophoresis, and Electroblotting (Step VII, Fig. 5)

The PCR-amplified fragments are separated by electrophoresis through an 8 % polyacrylamide/7 *M* urea gel, 0.4 mm thick and 60–65 cm long, and then transferred to a nylon membrane by electroblotting [11–13].

1. Add 25 µL of the appropriate stop mix: *Pfu* exo-, *Taq*, or *Vent* exo- stop mix, depending on the polymerase mix used for amplification.
2. Add 400 µL of precooled absolute ethanol and mix well.
3. Leave for 15 min on dry ice or at –80 °C until the samples are frozen and spin for 15 min at 16,000 × *g* at 4 °C.
4. Wash once with 500 µL of precooled 80 % ethanol.
5. Spin for 5 min at 16,000 × *g*.
6. Air-dry DNA pellets.
7. Dissolve DNA pellets in 7.5 µL of formamide loading dye in preparation for sequencing gel electrophoresis. For the sequence samples G, A, T + C, and C, it is often advisable to dissolve DNA pellets in 15 µL of formamide loading dye.
8. Prerun the 8 % polyacrylamide gel in 100 m*M* TBE, until the temperature of the gel reaches 50 °C.

9. To denature DNA, heat the samples at 95 °C for 2 min, and then keep them on ice prior to loading.
10. Wash the wells of the gel using a syringe.
11. Load an aliquot of 1.5–2 μL.
12. Run the gel at the voltage and power necessary to maintain the temperature of the gel at 50 °C. This will ensure that the DNA remains denatured.
13. Stop the gel when the green dye (xylene cyanol FF) reaches 1–2 cm from the bottom of the gel.
14. Separate the glass plates using a spatula, and then remove one of the plates by lifting it carefully. The gel should stick to the less treated plate (*see* **Note 12**).
15. Cover the lower part of the gel (approx 40–42 cm) with a clean Whatman 3MM Chr paper, carefully remove the gel from the glass plate, and cover it with a plastic film.
16. On the bottom plate of the electroblotter, individually layer three sheets of Whatman 17 MM Chr paper presoaked in 100 m*M* TBE and squeeze out the air bubbles between the paper layers by rolling with a bottle .
17. Add 150 mL of 100 m*M* TBE on the top layer and place the gel quickly on the Whatman 17 MM Chr papers before TBE is absorbed. Remove all air bubbles under the gel by gently rolling a 25 mL pipet.
18. Remove the plastic film and cover the gel with a positively charged nylon membrane presoaked in 100 m*M* TBE, remove all air bubbles by gently rolling a 25 mL pipet, then cover with three layers of presoaked Whatman 17 MM Chr paper, and squeeze out air bubbles with rolling bottle. Paper sheets can be reused several times.
19. Place the upper electrode onto the paper.
20. Electrotransfer for 45 min at 2 A. The voltage should be at approximately 10–15 V.
21. UV-cross-link (1000 J/m^2 of UVC) the blotted DNA to the membrane, taking care to expose the DNA side of the membrane. If probe stripping and rehybridization are planned, keep the membrane damp.

Preparation of Single-Stranded Hybridization Probes (Step VIII, Fig. 5)

The [^{32}P]-dCTP-labeled single-stranded probe is prepared by 30 cycles of repeated linear primer extension using *Taq* DNA polymerase. Primer 2 (or primer 3, *see* **Note 13**) is extended on a double-stranded template which can be a plasmid or a PCR product. The latter is produced by using two opposing primers 2 separated by a distance of 150–450 bp. Alternatively, any pair of gene specific primers suitable for amplifying a DNA fragment containing a suitable probe sequence (*see* **Note 13**) can be employed.

1. Prepare 150 µL of the isotopic labeling mix.
2. Cycle 30 times at 95 °C for 1 min (97 °C for 3 min for the first cycle), 60–68 °C for 2 min, and 74 °C for 3 min.
3. Transfer the mixture to a conical 1.5 mL microtube with screw cap containing 50 µL of 10 *M* ammonium acetate, 1 µL of glycogen, and 400 µL of precooled absolute ethanol.
4. Mix well, leave for 5 min at room temperature, and spin for 5 min at 16,000 × *g*.
5. Transfer the supernatant into a new 1.5 mL microtube. Using a Geiger counter, compare the counts per minute between the pellet (probe) and the supernatant, counts from the probe should be equal or superior to the counts from the supernatant for optimal results.
6. Dissolve the probe in 100 µL of TE buffer.
7. Add the probe to 6–8 mL of hybridization buffer and keep the probe at 65 °C.

Hybridization (Step VII, Fig. 5), Washing, and Autoradiography

1. Prehybridize the membrane with 20 mL of hybridization buffer at 65 °C for 20 min.
2. Decant the hybridization buffer and add the single-stranded hybridization probe in 6–8 mL of hybridization buffer.
3. Hybridize at 65 °C overnight.
4. Wash the membrane with prewarmed (65 °C) washing buffers. The membrane is placed into a tray on an orbital shaker. Wash with buffer I for 10 min and with buffer II three times for about 10 min each time.
5. Wrap the membrane in plastic film. Do not let the membrane become dry if stripping and rehybridization are planned after exposure of the film.
6. Expose membrane to X-ray films with intensifying screen at −80 °C. The exposure time depends of the cpm count evaluated with a Geiger counter. Nylon membranes can be rehybridized if more than one set of primers have been included in the primer extension and amplification reactions [11–13]. Probes can be stripped by soaking the membranes in boiling 0.1 % SDS solution twice for 5–10 min each time.

3.7.2 Fluorescent Method Using Sequencer

Labeling Extension

1. Add 3 µL of exonuclease I mix.
2. Incubate on PCR at 37 °C for 30 min and at 76 °C for 20 min.
3. Add 1 µL of primer 3.
4. Cycle five times as described in Table 5. The last extension should be done for 10 min to fully extend all DNA fragments.
5. Add 25 µL of labeling stop mix.

Table 5
PCR extension labeling program for *Taq* DNA polymerase

Cycles	Denaturation (T in °C for DF in s)	Annealing (T is the T_m of the oligonucleotide for D in s)	Polymerization (D in s) T is the same for all cycles: 74 °C
0	95 for 120		
1	95 for 45	T_m –2 °C for 120	180
Repeat cycle 1, 3 more times			
2	98 for 45	T_m –2 °C for 120	600

6. Add 300 µL of precooled absolute ethanol and mix well.
7. Leave for 15 min on dry ice or at –80 °C until the samples are frozen and spin for 15 min at 16,000 × *g* in a centrifuge at 4 °C.
8. Wash once with 500 µL of precooled 80 % ethanol.
9. Spin for 5 min at 16,000 × *g*.
10. Air-dry DNA pellets.
11. Dissolve DNA pellets in 15 µL of formamide loading dye in preparation for sequencing gel electrophoresis. For the sequence samples G, A, T + C, and C, it is often advisable to dissolve DNA pellets in 30 µL of formamide loading dye.
12. Incubate for 2 min at 95 °C and then keep on ice before loading on the sequencing gel.

Sequencer Electrophoresis

The labeled DNA fragments are separated by electrophoresis through an 8 % polyacrylamide/7 *M* urea gel, 0.2 mm thick and 66 cm long, using a LI-COR DNA 4300 sequencer (see **Note 14**).

1. Prerun the 8 % polyacrylamide gel 25 min in order to set the temperature at 47 °C. Running buffer is 100 m*M* TBE. Before loading the samples, wash the wells thoroughly using a syringe.
2. Load an aliquot of 1–1.5 µL.
3. Run the gel with a constant power of 100 W during 11 h.

4 Notes

1. Originally, *Pfu* exo- and *Taq* buffers were prepared using KCl which was, however, shown to stabilize secondary DNA structures, thus preventing an optimal polymerization [52]. The use of NaCl prevents, to some extent, the ability of DNA to form secondary structures. This is particularly helpful when GC-rich regions of the genome are being investigated.

2. Primers should be selected to have a higher T_m at the 5′ end than in the 3′ end. This higher annealing capacity of the 5′-end lowers false priming, thus allowing a more specific extension and less background [53]. A guanine or a cytosine residue should also occur at the 3′ end. This stabilizes the annealing and facilitates the initiation of the primer extension. It is important that the selected primer does have long runs of purines or pyrimidines, does not form loops or secondary structure, and does not anneal with itself. If primer dimerization occurs, less primer will be available for annealing and polymerization will not be optimal. The purity of the primers is verified on a 20 % polyacrylamide/7 M urea gel (to prepare 500 mL: dissolve 96.625 g acrylamide, 3.375 g *bis*-acrylamide, and 210.21 g urea in 100 m*M* TBE); if more than one band is found, the primer is reordered. The primers are also tested in a conventional PCR to prepare the template for the probe synthesis (*see* **Note 13**).
3. The genomic DNA used for LMPCR needs to be very clean and undegraded. Any shearing of the DNA during preparation and handling before the PE must be avoided. After an incubation of 3 h, if clumps of nuclei are still visible, proteinase K at a final concentration of 450 μg/mL should be added and the sample reincubated at 37 °C for another 3 h.
4. If no DNA can be seen, add glycogen (1–2 μg) to the DNA solution and put the DNA on dry ice or in −80 °C freezer until the samples are frozen and centrifuge the DNA (5000 × *g* for 20 min at 4 °C). This should help DNA recovery but increases the probability of RNA contamination.
5. Because in cellulo DNA analysis is based on comparison of DNA samples modified in cellulo with DNA control modified in vitro, given the quantitative characteristic and high sensitivity of LMPCR technology, the DNA concentrations should be as accurate as possible. Indeed, it is critical to start LMPCR with similar amounts of DNA in every sample to be analyzed. The method to evaluate DNA concentration should measure only double-stranded nucleic acids. RNA contamination does not affect LMPCR, although it can however interfere with the precise measurement of the DNA concentration.
6. The bottom of a capless 0.6 mL microtube can be easily pierced with a heated needle. It is important to emphasize that the hole should be made as small as possible for the column to efficiently retain agarose. The pierced microtube is packed with wetted glass wool. Three successive centrifugation steps of 1 min each at 16,000 × *g* are necessary to compact and dry the glass wool. The water is recovered in a capless 1.5 mL microtube. If glass wool is found with the effluent, the column should be discarded. A final 5 min centrifugation at 16,000 × *g*

should be carried out to ensure the glass wool is fully compacted and dry. The glass wool column is stored at room temperature in a new capless 1.5 mL microtube and covered with a plastic film to protect the column from dust. In this way, the column can be stored indefinitely until it is used.

7. The DNA break frequency is even more critical than the DNA concentration. For DMS and UV, the base-modification frequency determines the break frequency following conversion of the modified bases to single-strand breaks, whereas for DNase I, the frequency of cleavage is exactly the break frequency. The break frequency must be similar among the samples to be analyzed. It should not average more than one break per 150 bp for in cellulo DNA analysis, the optimal break frequency varying from one break per 200 bp to one break per 2000 bp. When the break frequency is too high, we typically observe dark bands over the bottom half of the autoradiogram and very pale bands over the upper half, reflecting the low number of long DNA fragments. In summary, to make the comparison of the in cellulo modified DNA sample with a in vitro DNA control easily interpretable and valid, the amount of DNA and the break frequency must be similar between the samples to be compared. On the other hand, it is not so critical that the break frequency of the sequence ladders (G, A, T + C, and C) be similar to that of the samples to be studied. However, to facilitate sequence reading, the break frequency should be similar between the sequence reactions. It is often necessary to load less DNA for the sequence ladders.
8. If the cell density is too high, multiple cell layers will be formed and the upper cell layer will obstruct the lower ones. This will result in an inhomogeneous DNA photoproduct frequency.
9. It is imperative that the in vitro DNA samples used as DNA control and the in cellulo samples come from the same cell type. For instance, differing cytosine methylation patterns of genomic DNA from different cell types affect photoproduct formation [2, 17] and give altered DNase I cleavage patterns [5].
10. A nearly ideal chromatin substrate can be maintained in permeabilized cells. Nonionic detergents such as lysolecithin [51] and Nonidet P40 Substitute [34] permeabilize the cell membrane sufficiently to allow the entry of DNase I. Conveniently, this assay can be performed with cells either in a suspension or in a monolayer. One concern is that permeabilized cells will lyse after a certain amount of time in a detergent; thus care must be taken to monitor cell integrity by microscopy during the course of the experiment. A further difficulty with the permeabilization technique concerns the relatively narrow detergent concentration range over which the assay can be [AU2]

performed. Each cell type appears to require specific conditions for the detergent cell permeabilization. Furthermore, the DNase I concentration must be calibrated for each cell type to produce an appropriate cleavage frequency. Optimally, the in cellulo DNase I protocol works better if the enzyme has cleaved the DNA backbone every 1.5–2 kb. Cutting frequencies greater than 1 kb are associated with higher LMPCR backgrounds because the number of 3′-OH ends is much higher, making the suppression of the extension of these ends more difficult.

11. So far with the method using the automated sequencer, we have tested and used only the *Pfu* exo- for primer extension and the *Taq* DNA polymerase for the PCR amplification.
12. To facilitate sequencing gel removal following migration, it is crucial to siliconize the inner face of both glass plates prior to pouring the gel. For security, cost effectiveness, efficiency and time-saving, we recommend to treat the glass plates with RAIN-AWAY™ solution (Wynn's Canada, product no. 63020). We apply 0.75 mL on one plate and 1.5 mL on the other before each utilization as specified by the manufacturer. In this way, the gel is easier to pour and will tend to stick on the less siliconized plate.
13. If a third primer (primer 3) is used to make the probe, it should be selected from the same strand as the amplification primer (primer 2), just 5′ to primer 2 sequence and with no or no more than seven to eight bases of overlap on this primer, and have a T_m of 60–68 °C. As first reported by Hornstra and Yang [43, 54, 55], we use the primer 2 employed in the amplification step and we produce the probe from PCR products. Such probes cost less (no primer 3) and are more convenient (the preparation of the PCR products permits the testing of primers).
14. To have well-shaped wells, we treat the upper part of both glass plates with a silane-acetic acid solution. We mix 50 µL of silane with 50 µL of 10 % acetic acid. Before each utilization, using a cotton-tipped applicator, we apply 100 µL of this solution on the plate areas where we place the comb.

Acknowledgments

This work was supported by the Canadian Genetic Diseases Network (MRC/NSERC NCE program) and the Canada Research Chair. R. Drouin holds the Canada Research Chair in "Genetics, Mutagenesis and Cancer."

References

1. Pfeifer GP, Tanguay RL, Steigerwald SD, Riggs AD (1990) In vivo footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. Genes Dev 4:1277–1287
2. Pfeifer GP, Drouin R, Riggs AD, Holmquist GP (1991) In vivo mapping of a DNA adduct at nucleotide resolution: detection of pyrimidine (6-4) pyrimidone photoproducts by ligation-mediated polymerase chain reaction. Proc Natl Acad Sci U S A 88:1374–1378
3. Chen CJ, Li LJ, Maruya A, Shively JE (1995) In vitro and in vivo footprint analysis of the promoter of carcinoembryonic antigen in colon carcinoma cells: effects of interferon gamma treatment. Cancer Res 55:3873–3882
4. Tornaletti S, Pfeifer GP (1995) UV light as a footprinting agent: modulation of UV-induced DNA damage by transcription factors bound at the promoters of three human genes. J Mol Biol 249:714–728
5. Pfeifer GP, Riggs AD (1991) Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permeabilized cells using DNase I and ligation-mediated PCR. Genes Dev 5:1102–1113
6. Mueller PR, Wold B (1989) In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. Science 246:780–786
7. Pfeifer GP, Steigerwald SD, Mueller PR, Wold B, Riggs AD (1989) Genomic sequencing and methylation analysis by ligation mediated PCR. Science 246:810–813
8. Pfeifer GP, Drouin R, Riggs AD, Holmquist GP (1992) Binding of transcription factors creates hot spots for UV photoproducts in vivo. Mol Cell Biol 12:1798–1804
9. Church GM, Gilbert W (1984) Genomic sequencing. Proc Natl Acad Sci U S A 81: 1991–1995
10. Pfeifer GP (1992) Analysis of chromatin structure by ligation-mediated PCR. PCR Methods Appl 2:107–111
11. Pfeifer GP, Riggs AD (1993) Genomic footprinting by ligation mediated polymerase chain reaction. In: White B (ed) PCR protocols: Current Methods and Applications. Humana, Totowa, NJ, pp 169–181
12. Pfeifer GP, Riggs AD (1993) Genomic sequencing, *Methods Mol.* Biol 23:169–181
13. Pfeifer GP, Singer-Sam J, Riggs AD (1993) Analysis of methylation and chromatin structure. Methods Enzymol 225:567–583
14. Gao S, Drouin R, Holmquist GP (1994) DNA repair rates mapped along the human PGK1 gene at nucleotide resolution. Science 263: 1438–1440
15. Tornaletti S, Pfeifer GP (1994) Slow repair of pyrimidine dimers at p53 mutation hotspots in skin cancer. Science 263:1436–1438
16. Rodriguez H, Drouin R, Holmquist GP, O'Connor TR, Boiteux S, Laval J, Doroshow JH, Akman SA (1995) Mapping of copper/hydrogen peroxide-induced DNA damage at nucleotide resolution in human genomic DNA by ligation-mediated polymerase chain reaction. J Biol Chem 270:17633–17640
17. Drouin R, Therrien JP (1997) UVB-induced cyclobutane pyrimidine dimer frequency correlates with skin cancer mutational hotspots in p53. Photochem Photobiol 66:719–726
18. Drouin R, Therrien JP, Angers M, Ouellet S (2001) In vivo DNA analysis. In: Moss T (ed) Methods in molecular Biology. Humana Press Inc, Totowa, NJ, pp 175–219
19. Dai SM, Chen HH, Chang C, Riggs AD, Flanagan SD (2000) Ligation-mediated PCR for quantitative in vivo footprinting. Nat Biotechnol 18:1108–1111
20. Dai SM, O'Connor TR, Holmquist GP, Riggs AD, Flanagan SD (2002) Ligation-mediated PCR: robotic liquid handling for DNA damage and repair. Biotechniques 33:1090–1097
21. Cartwright IL, Kelly SE (1991) Probing the nature of chromosomal DNA-protein contacts by in vivo footprinting. Biotechniques 11(188–190):192–184, 196 passim
22. Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol 65:499–560
23. Chin PL, Momand J, Pfeifer GP (1997) In vivo evidence for binding of p53 to consensus binding sites in the p21 and GADD45 genes in response to ionizing radiation. Oncogene 15:87–99
24. Angers M, Drouin R, Bachvarova M, Paradis I, Marceau F, Bachvarov DR (2000) In vivo protein-DNA interactions at the kinin B(1) receptor gene promoter: no modification on interleukin-1 beta or lipopolysaccharide induction. J Cell Biochem 78:278–296
25. Becker MM, Wang JC (1984) Use of light for footprinting DNA in vivo. Nature 309: 682–687
26. Pfeifer GP, Tornaletti S (1997) Footprinting with UV irradiation and LMPCR. Methods 11:189–196

27. Pfeifer GP, Chen HH, Komura J, Riggs AD (1999) Chromatin structure analysis by ligation-mediated and terminal transferase-mediated polymerase chain reaction. Methods Enzymol 304:548–571
28. Cadet J, Anselmino C, Douki T, Voituriez L (1992) Photochemistry of nucleic acids in cells. J Photochem Photobiol B 15:277–298
29. Mitchell DL, Nairn RS (1989) The biology of the (6-4) photoproduct. Photochem Photobiol 49:805–819
30. Holmquist GP, Gao S (1997) Somatic mutation theory, DNA repair rates, and the molecular epidemiology of p53 mutations. Mutat Res 386:69–101
31. Gale JM, Nissen KA, Smerdon MJ (1987) UV-induced formation of pyrimidine dimers in nucleosome core DNA is strongly modulated with a period of 10.3 bases. Proc Natl Acad Sci U S A 84:6644–6648
32. Gale JM, Smerdon MJ (1990) UV induced (6-4) photoproducts are distributed differently than cyclobutane dimers in nucleosomes. Photochem Photobiol 51:411–417
33. Mitchell DL, Nguyen TD, Cleaver JE (1990) Nonrandom induction of pyrimidine-pyrimidone (6-4) photoproducts in ultraviolet-irradiated human chromatin. J Biol Chem 265:5353–5356
34. Rigaud G, Roux J, Pictet R, Grange T (1991) In vivo footprinting of rat TAT gene: dynamic interplay between the glucocorticoid receptor and a liver-specific factor. Cell 67:977–986
35. Miller MR, Castellot JJ Jr, Pardee AB (1978) A permeable animal cell preparation for studying macromolecular synthesis. DNA synthesis and the role of deoxyribonucleotides in S phase initiation. Biochemistry 17:1073–1080
36. Contreras R, Fiers W (1981) Initiation of transcription by RNA polymerase II in permeable, SV40-infected or noninfected, CVI cells; evidence for multiple promoters of SV40 late transcription. Nucleic Acids Res 9:215–236
37. Tanguay RL, Pfeifer GP, Riggs AD (1990) PCR-aided DNaseI footprinting of single copy gene sequences in permeabilized cells. Nucleic Acids Res 18:5902
38. Tormanen VT, Swiderski PM, Kaplan BE, Pfeifer GP, Riggs AD (1992) Extension product capture improves genomic sequencing and DNase I footprinting by ligation-mediated PCR. Nucleic Acids Res 20:5487–5488
39. Rozek D, Pfeifer GP (1993) In vivo protein-DNA interactions at the c-jun promoter: preformed complexes mediate the UV response. Mol Cell Biol 13:5490–5499
40. Tornaletti S, Bates S, Pfeifer GP (1996) A high-resolution analysis of chromatin structure along p53 sequences. Mol Carcinog 17:192–201
41. Szabo PE, Pfeifer GP, Mann JR (1998) Characterization of novel parent-specific epigenetic modifications upstream of the imprinted mouse H19 gene. Mol Cell Biol 18: 6767–6776
42. Garrity PA, Wold BJ (1992) Effects of different DNA polymerases in ligation-mediated PCR: enhanced genomic sequencing and in vivo footprinting. Proc Natl Acad Sci U S A 89: 1021–1025
43. Hornstra IK, Yang TP (1994) High-resolution methylation analysis of the human hypoxanthine phosphoribosyltransferase gene 5′ region on the active and inactive X chromosomes: correlation with binding sites for transcription factors. Mol Cell Biol 14:1419–1430
44. Angers M, Cloutier JF, Castonguay A, Drouin R (2001) Optimal conditions to use Pfu exo(-) DNA polymerase for highly efficient ligation-mediated polymerase chain reaction protocols. Nucleic Acids Res 29, E83
45. Rouget R, Vigneault F, Codio C, Rochette C, Paradis I, Drouin R, Simard LR (2005) Characterization of the survival motor neuron (SMN) promoter provides evidence for complex combinatorial regulation in undifferentiated and differentiated P19 cells. Biochem J 385:433–443
46. Vigneault F, Drouin R (2005) Optimal conditions and specific characteristics of Vent exo-DNA polymerase in ligation-mediated polymerase chain reaction protocols. Biochem Cell Biol 83:147–165
47. Drouin R, Gao S, Holmquist GP (1996) Agarose gel electrophoresis for DNA damage analysis. In: Pfeifer GP (ed) Technologies for Detection of DNA Damage and Mutations. Plenum, New York, pp 37–43
48. Drouin R, Rodriguez H, Holmquist GP, Akman SA (1996) Ligation-mediated PCR for analysis of oxidative damage. In: Pfeifer GP (ed) Technologies for Detection of DNA Damage and Mutations. Plenum, New York, pp 211–225
49. Mueller PR, Wold B (1991) Ligation-mediated PCR: applications to genomic footprinting. Methods 20–31
50. Iverson BL, Dervan PB (1987) Adenine specific DNA chemical sequencing reaction. Nucleic Acids Res 15:7823–7830
51. Zhang L, Gralla JD (1989) In situ nucleoprotein structure at the SV40 major late promoter: melted and wrapped DNA flank the start site. Genes Dev 3:1814–1822

52. Fry M, Loeb LA (1994) The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. Proc Natl Acad Sci U S A 91:4950–4954
53. Rychlik W (1993) Selection of primers for polymerase chain reaction. In: White B (ed) PCR protocols: Current Methods and Applications. Humana, Totowa, NJ, pp 31–40
54. Hornstra IK, Yang TP (1992) Multiple in vivo footprints are specific to the active allele of the X-linked human hypoxanthine phosphoribosyltransferase gene 5′ region: implications for X chromosome inactivation. Mol Cell Biol 12:5345–5354
55. Hornstra IK, Yang TP (1993) In vivo footprinting and genomic sequencing by ligation-mediated PCR. Anal Biochem 213:179–193
56. McLellan JA (2001) Osmium tetroxide modification and the study of DNA-protein interactions. Methods Mol Biol 148:121–34

Chapter 5

Southwestern Blotting Assay

Yinshan Jia, Linda Nagore, and Harry Jarrett

Abstract

Southwestern blotting is a technique used to study DNA-protein interactions. This method detects specific DNA-binding proteins by incubating radiolabeled DNA with a gel blot, washing, and visualizing through autoradiography. A blot resulting from 1-dimensional SDS-PAGE reveals the molecular weight of the binding proteins. To increase separation and determine isoelectric point a 2-dimensional gel can be blotted. Additional dimensions of electrophoresis, such as a gel shift (EMSA), can precede isoelectric focusing and SDS-PAGE to further improve separation. Combined with other techniques, such as mass spectrometry, the DNA-binding protein can be identified.

Key words Transcription, Regulation, DNA, Response elements, Oligonucleotide, Promoter, Immunoblots, Nitrocellulose membrane, PVDF membrane, Southwestern blotting

1 Introduction

Regulation of gene expression is essential in human development as well as pathogenesis [1]. The gene regulatory mechanism involves distinct 6–8 bp DNA motifs referred to as response elements, which bind transcription factors [2]. Transcription factors are DNA-binding proteins that interact with unique response elements at the promoter region of DNA, or other functional cis-acting response elements, resulting in either gene expression or repression [1]. Identification of transcription factors specific to a particular gene is not only significant to gene regulation but also to understanding gene function. Southwestern blotting assay (SWB) is one of the most powerful techniques to explore protein-DNA interaction and transcription factor regulation.

SWB, similar to other blotting techniques, separates proteins (or DNA) by gel electrophoresis. The gel containing the separated proteins is electro-transferred (blotted) to a membrane, such as nitrocellulose (NC) or polyvinylidine difluoride (PVDF). To detect the DNA-binding proteins, the proteins are partially renatured and bound to nanomolar concentrations of radiolabeled DNA.

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_5,

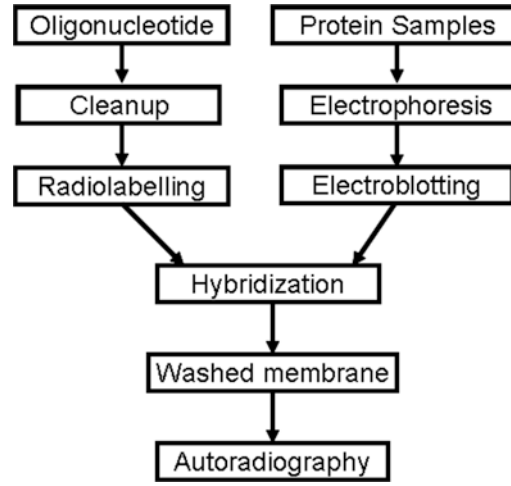


Fig. 1 Flow chart of SWB experimental procedure. Protein samples are either prepared from nuclear extract or purified by other techniques. Oligonucleotides are generally 20 bp in size containing one response element or using core promoter DNA

Any unbound DNA is washed away, and then bands on the blot are detected by autoradiography. This technique is especially useful in the identification of transcription factor as it gives information on the molecular weights of all DNA-binding proteins involved with a particular sequence of DNA. An experimental procedure is shown schematically in Fig. 1.

Since the original SWB assay, first described in 1980 [3], many extensions of the SWB method have been developed [4, 5]. In this chapter, three detailed SWB methods are described that differ in the number of electrophoresis dimensions used for protein separation. 1D-SWB uses separation on a single SDS-PAGE dimension, 2D-SWB uses separation by isoelectric focusing followed by SDS-PAGE in the second dimension, while 3D-SWB uses separation by DNA gel shift on a non-denaturing gel before 2D-SWB. Coupled with proteomic techniques, several transcription factors have been identified [4]. Additionally, a method has been developed that allows multiple reprobes with different oligonucleotides on one blot [6]. Each of these SWB techniques provides alternatives to investigate regulatory transcription factors in vitro.

2 Materials

2.1 Oligonucleotide Preparation

1. Oligonucleotide: Approximately 20–25 nt in size along with its complementary (antisense) strand, containing one protein-binding motif (0.1 μmol).

2. TE buffer: 10 mM Tris (free base), titrated to pH 7.5 with HCl, 1 mM EDTA.
3. 0.5 M EDTA (free acid): Titrated with 5 M NaOH to pH 8.0.
4. 3 M Sodium acetate solution: Dissolve 40.8 g sodium acetate ($C_2H_3NaO_2 \cdot 3H_2O$) in water to a final volume of 100 mL, and adjust to pH 5.2 with glacial acetic acid.

2.2 Oligonucleotide Labeling

1. ATP, [γ-^{32}P]: 6000 Ci/mmol, 10 μCi/μL, 250 μCi, stored at –20 °C.
2. T4 Polynucleotide Kinase: 10 U/μL, store at –20 °C.
3. 10 % Trichloroacetic acid (TCA): 10 g TCA dissolved to 100 mL in H_2O (*see* **Note 1**).
4. Silanized glass wool.
5. Culture tubes: Sterile, 17 × 100 mm sterile culture tubes.
6. Bio-Gel P-6 Fine Resin: Suspend 10 g of resin (Bio-Rad Laboratories, Hercules, CA, USA) in 250 mL of TE (pH 7.5) and autoclave for 45 min. Cool to room temperature, and then wash the resin with fivefold resin volumes TE three times. Remove excess liquid to give 1:1 slurry.

2.3 Cell Culture and Nuclear Protein Preparation

1. Human embryonic kidney 293 cells (HEK293).
2. Cell culture flasks: 182 cm^2.
3. Dulbecco's modification of Eagle's medium (DMEM).
4. Adult bovine serum: The serum is inactivated by incubation at 56 °C for 30 min and stored at 4 °C prior to use.
5. 10× PBS: 80 g NaCl, 2 g KCl, 14.4 g Na_2HPO_4, and 2.4 g KH_2PO_4 made up to 1 L with H_2O and autoclaved prior to use.
6. 1 M Dithiothreitol (DTT): Dissolve 15.4 g DTT in 100 mL of water. Aliquot and store at –20 °C for up to 6 months.
7. 1× Trypsin-EDTA.
8. Phenylmethylsulfonyl fluoride (PMSF, 0.2 M): Dissolve 3.48 g PMFS in 100 mL of anhydrous isopropanol as a 0.2 M stock. Stored at –20 °C for up to 1 year.
9. Nuclear extract hypotonic buffer: 10 mM HEPES, pH 7.9 at 4 °C, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM EDTA, 0.5 mM DTT (added prior to use), and 0.2 mM PMSF (added prior to use).
10. Nuclear extract low-salt buffer: 20 mM HEPES, pH 7.9 at 4 °C, 1.5 mM $MgCl_2$, 20 mM KCl, 0.2 mM EDTA, 25 % glycerol, 0.5 mM DTT, and 0.2 mM PMSF (both added prior to use).
11. Nuclear extract high-salt buffer: 20 mM HEPES, pH 7.9 at 4 °C, 1.5 mM $MgCl_2$, 1.6 M KCl, 0.2 mM EDTA, 25 %

glycerol, 0.5 mM DTT, and 0.2 mM PMSF (both added prior to use).

12. Nuclear extract dialysis buffer: 20 mM HEPES, pH 7.9 at 4 °C, 100 mM KCl, 20 % glycerol, 0.2 mM EDTA, 0.5 mM DTT, and 0.2 mM PMSF (both added prior to use).

2.4 Electroblotting and SWB assay

1. Electroblotting buffer: 20 % (v/v) methanol, 25 mM Tris-base, 192 mM glycine.
2. SWB buffer: 10 mM HEPES (pH 7.9 at 4 °C), 50 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 1 mM DTT, 50 μM $ZnSO_4$, and 0.1 % (v/v) Tween-20.
3. SWB blocking buffer: 5 % of nonfat dry milk in SWB buffer.
4. 6 M Guanidine hydrochloride: Dissolve 57.3 g guanidine hydrochloride (for molecular biology, ≥99 %) in 50 mL of SWB buffer, and then adjust to total 100 mL, freshly prepared prior to use.

2.5 1-Dimensional Gel Electrophoresis SWB (1DGE-SWB)

1. 30 % Acrylamide/Bis: Dissolve 29.2 g acrylamide and 0.8 g *N,N′*-methylene-bis-acrylamide (Bis) to a final volume of 100 mL H_2O.
2. 0.5 M Tris–HCl, pH 6.8: Dissolve 60.6 g Tris base in 800 mL H_2O, titrate to pH 6.8 with 1 M HCl, and adjust to 1 L with H_2O.
3. 1.5 M Tris–HCl, pH 8.8: Dissolve 181.7 g Tris base in 800 mL H_2O, titrate to pH 8.8 with concentrated HCl, and adjust to 1 L with H_2O.
4. 50 % Glycerol, 0.01 % bromophenol blue (BPB): To 50 mL glycerol, add 10 mg BPB, and adjust to 100 mL with H_2O.
5. 1 M Tris, pH 7.5: Dissolve 121.1 g Tris base in 800 mL H_2O. Titrate to pH 7.5 with concentrated HCl. Adjust volume to 1 L with H_2O and autoclave for 45 min.
6. 10 % (w/v) Ammonium persulfate (APS): Dissolve 100 mg APS (Sigma) in 1 mL H_2O, prepared prior to use.
7. β-Mercaptoethanol.
8. Tetramethylethylenediamine (TEMED).
9. 10 % (w/v) Sodium dodecylsulfate (SDS): Dissolve 10 g of SDS in H_2O to a final 100 mL; store at room temperature.
10. 5× Laemmli sample buffer: Mix 0.12 mL #2: 0.5 mL #4: 0.05 mL #7:0.2 mL # 9 on the day of use.
11. 10× Running buffer: 250 mM Tris base, 1.92 M glycine. Working solution is prepared with 30 mL 10× and 3 mL 10 % SDS and adjusted to 300 mL with H_2O and should be pH 8.3.

2.6 2-Dimensional Gel Electrophoresis SWB (2DGE-SWB)

1. Immobilized pH gradient (IPG) strip: (Bio-Rad) Ready@ Strip™ IPG strips, linear 7 cm, pH 3–10.
2. Rehydration buffer: 7 M Urea, 2 M thiourea, 2 % CHAPS, 65 mM DTT, 0.2 % Bio-Lyte 3/10 Ampholytes (Bio-Rad), 1 % Zwittergent 3–10 (Sigma), and 0.001 % bromophenol blue.
3. Equilibration buffer (EB): 50 mM Tris, pH 8.8, 6 M urea, 2 % (w/v) SDS, 30 % (v/v) glycerol, and 0.001 % (w/v) bromophenol blue.
4. Reduction buffer: 2 % (w/v) DTT in EB.
5. Alkylation buffer: 2.5 % (w/v) Iodoacetamide in EB.

2.7 EMSA-Based SWB (3DGE-SWB)

1. 5× EMSA buffer: 100 mM HEPES (pH 7.9), 0.5 mM EDTA, 250 mM NaCl, 25 mM $MgCl_2$, 5 mM DTT, 50 % (v/v) glycerol, and 0.5 % (v/v) Tween-20. Store at –20 °C for 1 month.
2. Poly (deoxyinosinic-deoxycytidylic) acid (poly-dI:dC): (Sigma) Dissolved in TE to bring to 0.5 mg/mL stock. Stored at –20 °C for at least 3 months.
3. 5× Loading buffer: 50 % (v/v) glycerol, 0.01 % (w/v) bromophenol blue.
4. 5× TBE buffer: 54.5 g Tris base, 27.8 g boric acid, 3.7 g Na_4EDTA, pH 8.3 at room temperature.
5. Extraction buffer: 50 mM Tris, pH 9.0, 50 mM DTT, and 0.5 % (v/v) Tween-20.

3 Methods

3.1 Oligonucleotide Preparation

1. Dissolve received oligonucleotide in 300 μL of TE (pH 7.5) buffer (*see* **Note 2**).
2. Add 30 μL of sodium acetate solution (3 M) and 1 mL of 100 % ethanol (absolute), mix, and allow oligonucleotides to precipitate at –85 °C for 1 h (*see* **Note 3**).
3. Collect oligonucleotides by centrifugation at 14,000 × *g* using bench centrifuge for 15 min at 4 °C (*see* **Note 4**).
4. Carefully remove the supernatant and wash the pellet with 500 μL of ice-cold 70 % (v/v) ethanol by vortex mixing.
5. Again, centrifuge and remove the supernatant as in **step 4**.
6. Air-dry the oligonucleotide by leaving the tube uncapped and covered with Kimwipes for 2 h at room temperature.
7. Dissolve the pellet in 500 μL of TE (pH 7.5) at room temperature for 30 min with occasional vortex mixing.

8. Determine the concentration by measuring the absorption at 260 nm using 1 μL oligonucleotide diluted to 1 mL water (*see* **Note 5**).
9. Adjust the concentration to 0.1 mM with TE (pH 7.5).

3.2 Oligonucleotide Labeling

1. Caution: All steps must be carried out in a radioactivity control area.
2. Mix 2 μL of 10 μM oligonucleotide, 5 μL of polynucleotide kinase buffer (10×, supplied with enzyme), 2 μL of γ-^{32}P-ATP (10–20 μCi), and 2 μL of T4 Polynucleotide Kinase (10 U/μL), and bring to 50 μL with water in a 1.5 mL Eppendorf tube. Mix by gentle tapping; centrifuge briefly.
3. Incubate at 37 °C for 60 min.
4. Stop reaction by adding 2 μL of 0.5 M EDTA.
   To carry out TCA analysis remove 1 μL reaction mixture from **step 4** to a 13×100 mm test tube containing 100 μL of 100 μg/mL salmon sperm DNA in TE. Mix well.

   Spot 1 μL of this mixture directly onto a Whatman GF/C 47 mm filter disk.

   To the remaining 100 μL, add 5 mL of ice-cold 10 % TCA, vortex, and leave on ice for 15 min.

   Collect precipitate by vacuum filtration through a GF/C filter. Wash the tube and filter five times with 5 mL of ice-cold TCA, and then twice with 5 mL of ice-cold ethanol.
5. Count both filters by Cherenkov radiation (without) or with 5 mL of scintillation fluid. Total counts are obtained from the directly spotted 1 μL (**step 6**) by:

   $$\text{Total} = \text{C.P.M.} \times (101\ \mu\text{L}/1\ \mu\text{L}) \times (52\ \mu\text{L}/1\ \mu\text{L})$$

   And that precipitated (**step 8**):

   $$\text{TCA} = \text{C.P.M.} \times (100\ \mu\text{L}/101\ \mu\text{L}) \times (52\ \mu\text{L}/1\ \mu\text{L})$$

   $$\text{Labeling efficiency} = [\text{CPM (from TCA)}/\text{CPM (Total)}] \times 100\,\%$$
6. The oligonucleotide is desalted on a spin column (commercially available). The column is centrifuged for 5 min at 2000×*g* with a countertop centrifuge. The concentration of the oligonucleotide is now approximately 400 nM (*see* **Note 6**).
7. Equal amount of sense and antisense oligonucleotides are mixed and annealed by heating to 95 °C for 5 min, and then slowly cooling to room temperature over 1 h, or using a thermo-cycler annealing program.
8. The specific activity of the oligonucleotide is adjusted to 4,000,000 cpm/mL and 10 nM labeled oligonucleotide with TE buffer and/or unlabeled oligonucleotide, and then stored in 50 μL of aliquots at −20 °C (*see* **Note 7**).

3.3 Nuclear Extract Preparation 7

1. Culture human embryonic kidney 293 (HEK293) cells in a NuAire IR Autoflow CO_2 water-jacketed incubator at 37 °C with 5 % CO_2 and 95 % atmospheric air. Seed 182 cm^2 cell culture flasks with 5×10^6 HEK293 cells and grow in 60 mL/flask of DMEM containing 10 % heat-inactivated adult bovine serum. For a typical preparation, we grow 5–10 flasks.
2. Grow cells to 90 % confluence.
3. Rinse the cells with 1× PBS and then harvest by adding 10 mL of trypsin-EDTA solution for 5 min at 37 °C.
4. Immediately add 10 mL of DMEM containing 10 % inactivated adult bovine serum to stop trypsin.
5. Remove the cells from the flask and suspend in 40 mL 4 °C DMEM containing 10 % inactivated adult bovine serum. This method is a minor modification of the method of Abmayr et al. [7]. Keep all subsequent steps at 4 °C.
6. Centrifuge the cells (1850×*g* for 5 min, 4 °C) in 50 mL disposable, sterile plastic conical tubes and wash with 50 mL PBS. Carefully remove the supernatant and note the packed cell volume (pcv) (*see* **Note 8**).
7. Quickly resuspend the cells in five pcv of ice-cold nuclear extract hypotonic buffer, centrifuge the cells (1850×*g* for 5 min, 4 °C), and discard the supernatant.
8. Resuspend the cell pellet in three pcv of nuclear extract hypotonic buffer and swell for 10 min on ice.
9. Transfer the cells to an ice-cold Dounce homogenizer and homogenize with the type B pestle using ten slow up-and-down strokes.
10. Collect the nuclei by centrifugation (3300×*g*, 15 min, and 4 °C) in a graduated conical tube. Discard the supernatant and note the packed nuclear volume (pnv).
11. Add half pnv of nuclear extract low-salt buffer. Gently mix, and add dropwise half pnv of nuclear extract high-salt buffer, gently stirring for 30 min on ice.
12. Transfer the extracted nuclei to a JA-20 centrifuge tube and centrifuge at high speed (25,000×*g*) for 30 min. Save the supernatant and discard the pellet.
13. Dialyze the supernatant nuclear extract three times versus 50 volumes of nuclear extract dialysis buffer, allowing 4 h between each buffer change, 12 h total (*see* **Note 9**).
14. Determine the protein concentration using Bradford protein assay and dilute as needed with fresh dialysis buffer for a final concentration of 5 mg/mL protein (approximately 1–2 mL final volume).

15. Centrifuge the dialyzed nuclear extract (25,000×*g*, 20 min), carefully aliquot 50–100 μL of the supernatant into cooled 1.7 mL Eppendorf tubes, and store at −85 °C for up to 1 year (*see* **Note 10**).

3.4 1-Dimensional Gel Electrophoresis SWB (1DGE-SW)

1. Pour 12 % polyacrylamide mini-gel with a Bio-Rad Mini-PROTEIN empty cassette gel casting system, as per the manufacturer's protocol (*see* **Note 11**).
2. Mix 20 μL of protein sample and 5 μL of Laemmli buffer (5×) and heat at 95 °C for 5 min to denature proteins.
3. Load 25 μL of sample in each well plus one well for the molecular mass markers (*see* **Note 12**).
4. Fill electrophoresis cell with running buffer (1× containing 0.1 % SDS) and electrophorese at 100 V for ~1.5 h at room temperature until the BPB tracking dye is within 2–3 mm from the bottom.
5. Electroblot protein on the gel to PVDF membrane at 4 °C for 2 h with 100 mA in electroblotting buffer, using Bio-Rad electroblotting system.
6. Incubate blot in 25 mL of 6 M of guanidine solution for denaturing for 10 min.
7. Renaturing is by serially diluting the guanidine HCl solution with SWB buffer to a final concentration of 3, 1.5, 0.75, 0.375, 0.188, and 0.094 M. Between each dilution incubate the blot for 10 min at 4 °C (*see* **Note 13**).
8. Wash blot twice with 25 mL of SWB buffer, and place blot in 50 mL of SWB blocking buffer for 1 h at room temperature with gentle rocking (*see* **Note 14**).
9. During **step 8**, prepare SWB buffer containing [γ-^{32}P]-ATP radiolabeled oligonucleotide (1.5–10 nM, 10^6 cpm/mL), 0.25 % BSA, and 10 μg/mL poly dI:dC (*see* **Note 15**).
10. Place blot from **step 8** in the solution prepared in **step 9** and incubate at 4 °C with gentle rocking overnight.
11. Wash blot with 50 mL of SWB buffer at least three times (*see* **Note 16**).
12. Air-dry washed blot, and then expose to film for autoradiography. A 1DGE-SWB is shown in Fig. 2a (*see* **Note 17**).

3.5 2-Dimensional Gel Electrophoresis SWB (2DGE-SWB)

1. Mix 25 μL of protein sample (50–150 μg NE) with 100 μL IEF rehydration buffer, and distribute the sample mixture evenly across one well of the focusing tray using Bio-Rad PROTEIN IEF system (*see* **Note 18**).
2. Carefully remove the IPG strip (7 cm, linear, pH 3–10) protective cover with forceps starting from the acidic end (pH 3) and pulling towards the basic end (pH 10).

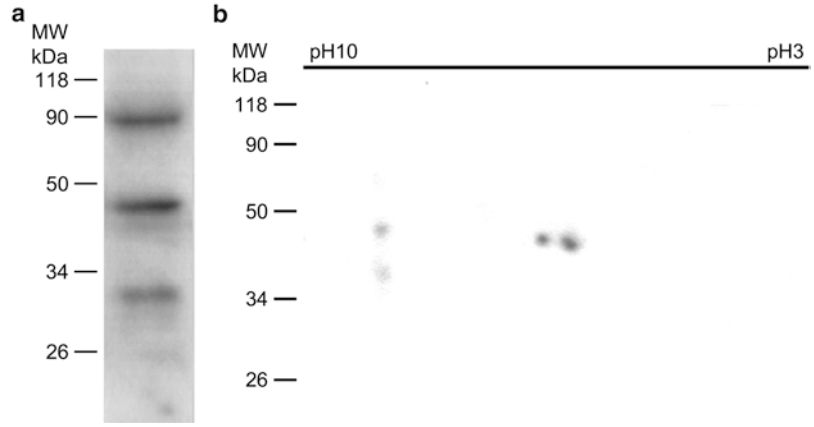


Fig. 2 1DGE-SWB and 2DGE-SWB results. Protein sample separated by an SDS-PAGE gel alone is called 1DGE-SWB. (**a**) A lane of a 12 % SDS-PAGE containing 1 μg of HEK293 NE was electroblotted onto a PVDF membrane and probed with 10 nM of an oligonucleotide containing the response element bound by MEF2 (5′-TGGGCTATTTTTAGGGG-3′, annealed to its complement strand). (**b**) NE (50 μg) was separated by IEF and 12 % SDS-PAGE and electroblotted to PVDF. It was then probed at 4 °C with 10 nM of the MEF2 oligonucleotide

3. Place the gel side on the IPG strip down on top of the sample mixture with the acid end (pH 3) of the strip on the positive electrode in the tray. Overlay strip with mineral oil to prevent evaporation.
4. Perform active rehydration mode at 50 V for 12 h (*see* **Note 19**).
5. Place paper wicks at each end of strip between strip and electrode. Focus at 20 °C with 40,000 V · h.
6. Remove focused strip from tray and wash with rehydration buffer to remove mineral oil.
7. Place IPG strip in 15 mL of reduction buffer in a 15 mL conical tube. Incubate at room temperature with rocking for 15 min.
8. Transfer strip into another 15 mL conical tube containing 15 mL of alkylation buffer; incubate at room temperature in the dark with rocking for 15 min.
9. Rinse strip with 50 mL of water at room temperature, and then place onto a 12 % polyacrylamide hand-casting mini gel (*see* **Note 20**), as well as a protein ladder.
10. During **step 8**, heat 1 % agarose sealing solution, containing 0.001 % (w/v) of bromophenol blue and 125 mM Tris (pH 6.8), at 90 °C for 5 min to liquefy, and then cool to 50 °C before sealing. Apply agarose over the strip, carefully avoiding air bubbles.
11. Run SDS-PAGE at 100 V for 1.5 h or until the blue dye almost reaches the bottom of the gel.

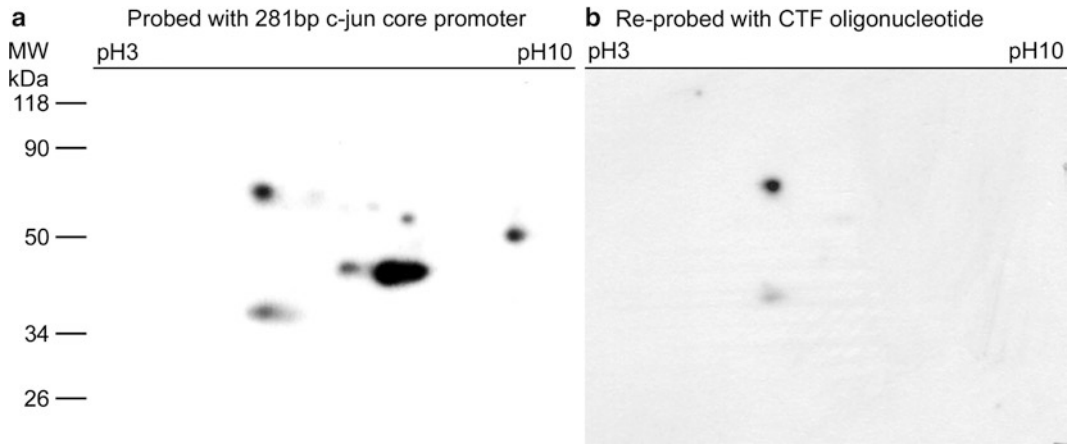


Fig. 3 Repeatedly probed SWB assay after enzymatic stripping. (**a**) 50 μg HEK293 NE was separated by 2DGE. Proteins were electroblotted onto a PVDF membrane and SWB assay was performed as described in Subheading 3.5. The blot was first probed with 10 nM radiolabeled core c-jun promoter DNA (281 bp). (**b**) The blot was then stripped by alkaline phosphatase (2 U/mL) as described in [6] and reprobed with 10 nM of a radiolabeled oligonucleotide (5′-ACGCGAGCCAATGGGAAG-3, annealed with its complement) bound by CTF

12. Remove the gel from running chamber, and place into a clean plastic tray for blotting.
13. Repeat **steps 5–12** in Subheading 3.4. A 2DGE-SWB is shown in Fig. 2b; the results of probing a 2DGE-SWB with one probe, stripping [6], and reprobing with a different probe are shown in Fig. 3.

3.6 EMSA-Based 3D-SWB (3DGE-SWB)

1. Incubate 10 μL nuclear extract (50 μg) with radiolabeled 1.6 nM core promoter DNA or duplex oligonucleotide to form protein-DNA complex in EMSA buffer containing poly dI:dC (20 ng/μL) in a total volume of 25 μL at room temperature for 20 min and on ice for 10 min.
2. The 3 % or 6 % non-denaturing polyacrylamide gels for promoter complex or oligonucleotide complex, respectively, are pre-run at 4 °C for 30 min at 50 V.
3. The protein mixture is applied to the gel for electrophoresis for 90 min at room temperature at 90 V (*see* **Note 21**).
4. The gel is electrotransferred to a PVDF membrane at 4 °C for 1.5 h in 20 % (v/v) methanol, 25 mM Tris base, and 192 mM glycine buffer at 100 mA.
5. After autoradiography, the DNA-protein complex is cut from the PVDF membrane at the position containing the specific complex using a clean, sharp scalpel. Then, the blotted band is cut into 1 × 1 mm pieces.
6. Place pieces into a 1.7 mL Eppendorf tube. Extract pieces twice with 1 mL of extraction buffer to soak the PVDF membrane

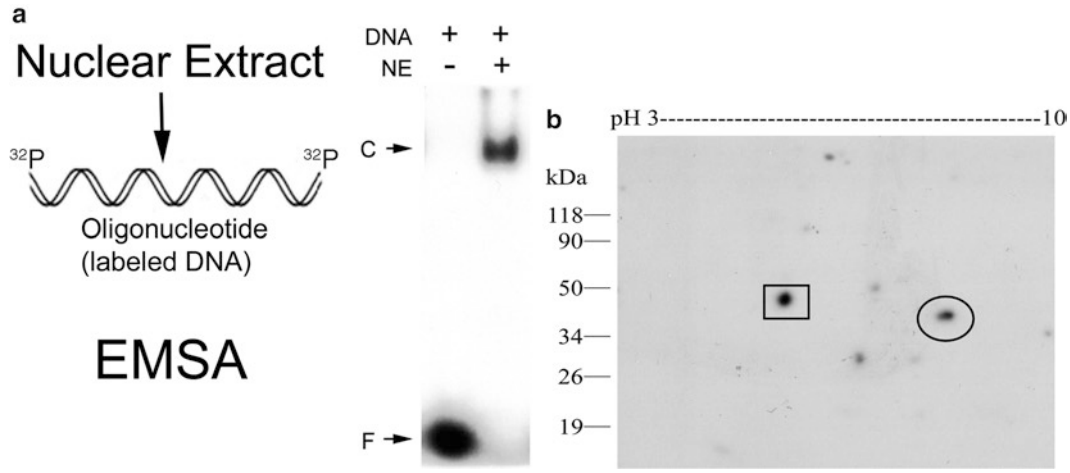


Fig. 4 3DGE-SWB using an AP1 oligonucleotide. (**a**) Fifty micrograms of HEK293 nuclear extract was mixed with 1.6 nM radiolabeled AP1 oligonucleotide (5′-CGCTTGATGACTCAGCCGGAA-3′ annealed with its complement), separated by 5 % non-denaturing PAGE, and electroblotted to PVDF. After autoradiography, the complex (C) band was excised and the proteins extracted. (**b**) The extract was then applied to 2DGE-SWB probed with 10 nM radiolabeled AP1 oligonucleotide. Panel **b** is reprinted from J. Chromatgr. A 1218, Jiang, D., Jia, Y., and Jarrett, H.W. "Transcription factor proteomics: Identification by a novel gel mobility shift-three-dimensional electrophoresis method coupled with southwestern blot and high-performance liquid chromatography–electrospray-mass spectrometry analysis", pages 7003–7015 (2011) with permission from Elsevier

and elute proteins at room temperature for 2 h each with gentle rocking. Combine the extracts.

7. Remove detergent from the extract using a Pierce Detergent Removal Spin column according to the manufacturer's protocol.
8. Concentrate protein sample to 20–25 μL using an Amicon Ultra 0.5 mL centrifugal concentrator (10,000 molecular mass cutoff) for IEF.
9. Repeat Subheading 3.5 for 2DGE-SWB. A 3DGE-SWB is shown in Fig. 4.

4 Notes

1. Water is >18 MΩ from a Millipore Synergy UV water purification unit.
2. Oligonucleotides are purchased from Integrated DNA Technologies (Coralville, IA, USA). Molar absorptivity (E_{260nm}) is provided to calculate concentration.
3. For short oligonucleotides (<15 nt.), increasing ethanol to 1.2 mL and using 75 % ethanol for the washes may increase yield.

4. Observe the tube orientation so that the position of the pelleted DNA, which may not be visible, is known and can be avoided.
5. Alternatively, a NanoDrop spectrophotometer can be used; however the results may be less accurate.
   (a) The oligonucleotide can be desalted on a commercially available column or they can be made in-house as shown in Fig. 5. Start by plugging the outlet of a 1 mL tuberculin syringe barrel with silanized glass wool.
   (b) Fill the syringe barrel with 1 mL of 1:1 slurry of Bio-Gel P-6 resin.
   (c) The column is placed inside a 17 × 100 mm culture tube with a hole punctured through the lid (scissors work well) to collect eluate from the column.
   (d) Centrifuge the column at 2000 × *g* for 5 min. If unsure about the time, repeat centrifugation for additional time until no further liquid elutes.

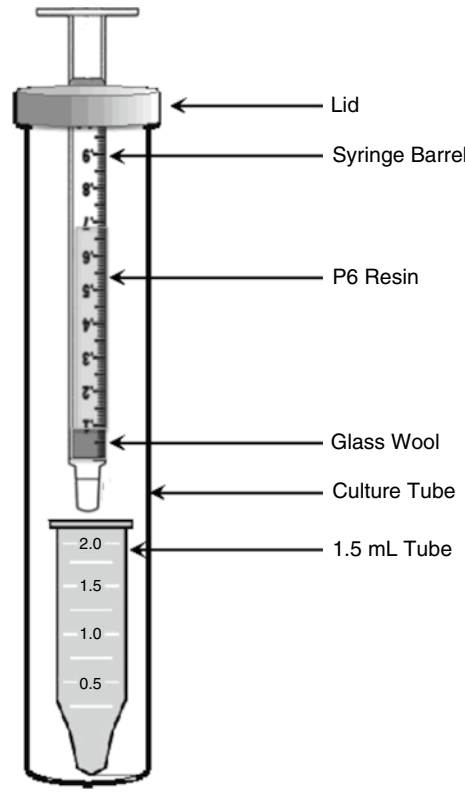


Fig. 5 A simple homemade desalting spin column. The column was prepared as described in **Note 6**

(e) Discard the eluate and add the labeled DNA to the column. Place a 1.5 mL Eppendorf tube (without lid) under the column and replace in the 17 × 100 mm carrier tube.

(f) Centrifuge the column and collect the eluate containing labeled DNA into a clean 1.5 mL centrifuge tube.

6. Desalting removes over 90 % of the remaining γ-^{32}P-ATP. The Eppendorf tube used for collection can be pre-weighed and weighed again after collection to determine the volume more accurately.
7. If labeling efficiency (Subheading 3.2, **step 5**) is 50 %, after desalting the DNA will be 50 μL of 400 nM oligonucleotide, and contain 20×10^6 cpm. In this case, the 50 μL labeled oligonucleotide would be diluted to 2 mL with TE, resulting in 10 nM oligonucleotide. To this would be added 3 mL 10 nM unlabeled oligonucleotide (in TE) to give a total of 5 mL and 4×10^6 cpm/mL. Generally, enough 50 μL aliquots are prepared for 2 weeks of experiments and the rest discarded in liquid waste.
8. The pcv is approximately 4 mL (from ten of 182 cm^2 cell culture flasks) and contains 4×10^8 cells.
9. Long dialysis is detrimental to activity. The dialysis procedure described preserves activity and thorough dialysis.
10. Nuclear extract can be stored for at least 1 year at –85 °C.
11. The percentage of SDS-PAGE used is dependent on the molecular weight of the protein of interest. For two 12 % polyacrylamide SDS-PAGE mini-gels mix 4 mL of 30 % acrylamide/BIS solution, 2.5 mL of 1.5 M Tris–HCl, pH 8.8, 5 μL of TEMED, 50 μL of 10 % APS, and 3.35 mL of water for separating gel. Add to the gel cassette to 1.5 cm from the top and overlay with H_2O. The gel will polymerize in approximately 20 min, after which the H_2O layer is removed. For the upper (stacking) gel: 1.3 mL of 30 % acrylamide/BIS solution, 2.5 mL of 0.5 M Tris base pH 6.8, 10 μL of TEMED, 50 μL of APS, and 6.1 mL of H_2O. Add to the cassette containing the well comb and fill completely. The upper gel will require 30–60 min for polymerization.
12. Load 10 μL of 1× Laemmli buffer into any unused wells to maintain even charge distribution.
13. Protein renaturing is a crucial step in SWB assay. Some proteins may be difficult to renature from a denatured structure; consequently adjustment of the temperature and/or time may be needed for optimization.
14. The blocking procedure chosen is dependent on subsequent experiments. For example, the blot can be blocked with 0.5 % polyvinyl pyrrolidone (PVP-40) in SW buffer; however,

PVP-40 fragments cause complex spectra in a mass spectrometer. Nonetheless PVP-40 works best for renaturation and retention on the blot and gives the highest signal for SWB. Also, bovine serum albumin is not advised by most commercial mass spectrometry facilities as it may cause peak suppression in the proteins of interest. Alternatively, overnight blocking is recommended for better renaturing protein on the blot.

15. Poly dI:dC is a commonly used competitor that reduces nonspecific interactions. Other competitors, such as mutant oligonucleotides or single-stranded DNA may also be applied to increase specific binding results.
16. Efficient washing steps will result in increasing the signal-to-noise ratio as well as reduce nonspecific interactions.
17. Exposure time can be adjusted to increase or decrease the strength of the signal.
18. Enhanced rehydration buffer is used here. The standard rehydration buffer includes 8 M urea and eliminated thiourea.
19. Passive rehydration is an alternative mode to rehydrate protein.
20. SDS-PAGE gel prepared with one long well. Two ends of IPG strip are cut off about half cm and inserted into the long well including a piece of filter paper containing absorbed 5 μL of protein ladder at one end. The top of strip is sealed with 1 % agarose sealing solution.
21. The percentage of acrylamide is dependent on the size of the DNA-protein complex to be separated. Optimal percentages range from 3 to 4 % for large complex, such as TF-core promoter DNA complexes, and 5–7 % for short oligonucleotide (20–25 bp) with NE.

Acknowledgement

This work was supported by NIH grant R01GM043609. LIN receives partial stipend support from NIH/NIGMS MBRS-RISE GM060655.

References

1. Lee TI, Young RA (2013) Transcriptional Regulation and Its Misregulation in Disease. Cell 152:1237–1251
2. Hutchins AP, Diez D, Takahashi Y, Ahmad S, Jauch R, Tremblay ML, Miranda-Saavedra D (2013) Distinct transcriptional regulatory modules underlie STAT3's cell type-independent and cell type-specific functions. Nucleic Acids Res 41:2155–2170
3. Bowen B, Steinberg J, Laemmli UK (1980) The detection of DNA-binding proteins by protein blotting. Nucleic Acids Res 8:1–20
4. Jiang D, Jia Y, Jarrett HW (2011) Transcription factor proteomics: identification by a novel gel

mobility shift-three-dimensional electrophoresis method coupled with southwestern blot and high-performance liquid chromatography-electrospray-mass spectrometry analysis. J Chromatogr A 1218:7003–7015

5. Jiang D, Jia Y, Zhou Y, Jarrett HW (2009) Two-dimensional southwestern blotting and characterization of transcription factors on-blot. J Proteome Res 8:3693–3701
6. Jia Y, Jiang D, Jarrett HW (2010) Repeated probing of Southwestern blots using alkaline phosphatase stripping. J Chromatogr A 1217: 7177–7181
7. Abmayr SM, Yao T, Parmley T, Workman J (2006) Preparation on nuclear and cytoplasmic extracts from mammalian cells. Curr Protoc Mol Biol Chapter 12:Unit 12.1. doi: 10.1002/0471142727.mb1201s75

Chapter 6

Single-Molecule Approaches for the Characterization of Riboswitch Folding Mechanisms

Julien Boudreault, D. Cibran Perez-Gonzalez, J. Carlos Penedo, and Daniel A. Lafontaine

Abstract

Riboswitches are highly structured RNA molecules that control genetic expression by altering their structure as a function of metabolite binding. Accumulating evidence suggests that riboswitch structures are highly dynamic and perform conformational exchange between structural states that are important for the outcome of genetic regulation. To understand how ligand binding influences the folding of riboswitches, it is important to monitor in real time the riboswitch folding pathway as a function of experimental conditions. Single-molecule FRET (sm-FRET) is unique among biophysical techniques to study riboswitch conformational changes as it allows to both monitor steady-state populations of riboswitch conformers and associated interconversion dynamics. Since FRET fluorophores can be attached to virtually any nucleotide position, FRET assays can be adapted to monitor specific conformational changes, thus enabling to deduce complex riboswitch folding pathways. Herein, we show how to employ sm-FRET to study the folding pathway of the *S*-adenosylmethionine (SAM) and how this can be used to understand very specific conformational changes that are at the heart of riboswitch regulation mechanism.

Key words Riboswitch, RNA folding, Single-molecule FRET, Metabolite-sensing, *S*-adenosylmethionine

1 Introduction

Riboswitches are highly structured domains that are found in mRNA untranslated regions. These regulatory switches control genetic expression by changing their structure as a function of metabolite binding. Riboswitches are composed of an aptamer domain that binds to a metabolite and an expression platform that modulates genetic expression. The aptamer is highly conserved and ensures high ligand binding affinity and specificity. The expression platform is highly variable and controls genetic expression at the level of transcription, translation, mRNA splicing, and mRNA decay [1, 2]. The ligand-dependent riboswitch structural

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_6,

reorganization is key in the regulatory process and it is thus vital to characterize how these structural conformers are contributing to the genetic regulation.

FRET is unique in its ability to monitor structural changes occurring in biological molecules since it can provide long-range distance information (20–80 Å) [3]. The application of FRET at the single-molecule level enables to characterize nucleic acid dynamics, which is generally hidden in bulk assays [4]. Furthermore, sm-FRET is useful to identify subpopulations in a heterogeneous mixture, which is essential to establish the folding pathway and dynamics of biomolecules. Sm-FRET was recently used to characterize the folding pathway of various riboswitches [5–10], providing a wealth of information regarding how ligand-dependent structural reorganization is used to control genetic expression. Clearly, based on the diversity of ligand binding and folding pathways that riboswitches employ to attain the native state, it suggests that a wide array of folding mechanisms are harnessed by riboswitches to control genetic expression.

Sm-FRET was previously employed to establish that ligand recognition by *S*-adenosylmethionine (SAM-I) riboswitch involves key conformational changes in the RNA architecture [6]. The aptamer domain contains a four-way helical junction (P1–P4 stems) that undergoes a two-step hierarchical folding induced by metal ions and SAM binding (*see* Fig. 1). The binding of magnesium ions promotes the formation of a structure (F_{Mg}) that corresponds to the close juxtaposition of stems P1 and P3, the stacking of stems P2 and P3, as well as the formation of the P3-J3/4 pseudoknot interaction. Ligand binding further promotes an additional conformer (F_{NS}) in which a helical stack involving stems P1 and P4 is formed. Ligand binding also promotes the helical rotation of the P1 stem along its axis, which is likely to be central for the specific

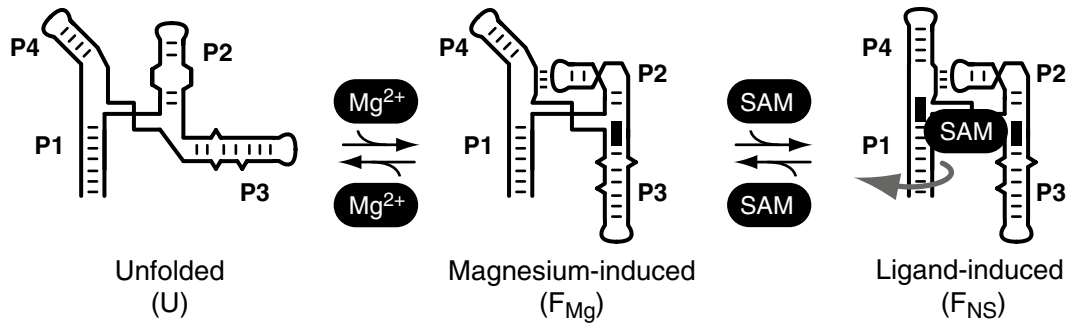


Fig. 1 Folding pathway of the SAM-I riboswitch aptamer. Conformational transitions occurring between the unfolded (U), magnesium-induced (F_{Mg}), and SAM-induced (F_{NS}) structures are shown. The *curved arrow* represents the helical rotation of the P1 stem in the F_{NS} state. The nomenclature for each helical domain is shown for each state

recognition of ligand. Herein, we provide detailed information about the use of sm-FRET to study the folding of the SAM-I riboswitch aptamer as a function of experimental conditions.

2 Materials

2.1 Assembly of the Riboswitch SAM-I Aptamer

The aptamer domain is reconstituted from a mixture of transcribed and synthetic RNA strands (all sequences are written in the 5′ to 3′ direction) (*see* Fig. 2a). In this construct, Cy3 donor and Cy5 acceptor fluorophores are attached via amino linkers (5NU) to strands P3 and P4–P1, respectively. Since fluorophore labeling and methods for aptamer reconstitution have been amply described previously [11, 12], we will only provide sequences required to obtain a SAM-I aptamer having fluorophores on stems P1 and P3.

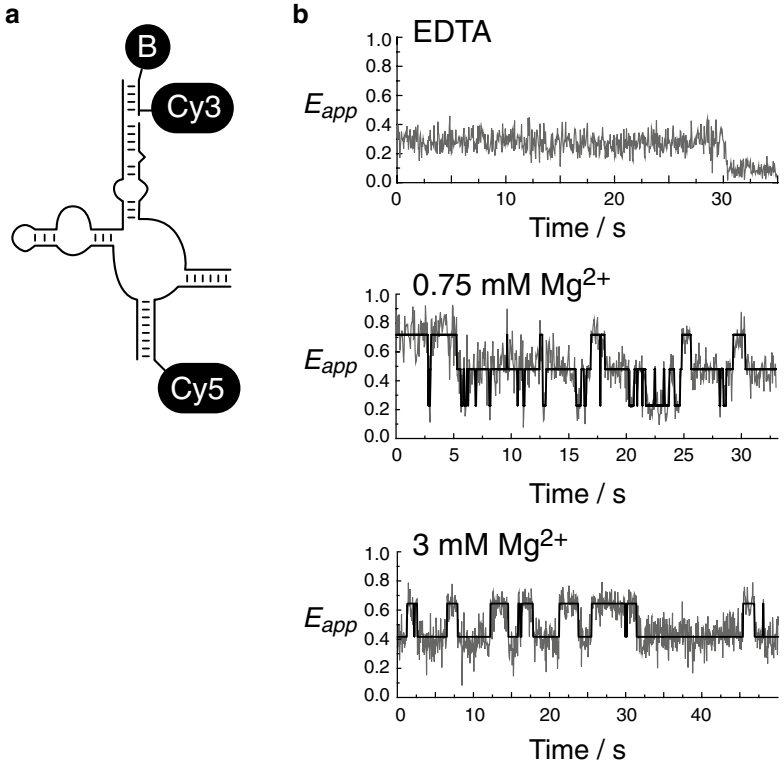


Fig. 2 Single-molecule FRET of the P1–P3 SAM-I construct. (**a**) The P1–P3 FRET construct used for single-molecule FRET experiments. The construct is reconstituted by assembling four RNA strands corresponding to the SAM-I aptamer. A biotin is present in the P3 stem to allow binding on the mounted channel imaged during sm-FRET analysis. (**b**) Single-molecule FRET histograms obtained in the presence of 50 mM EDTA, 0.75 mM Mg^{2+}, or 3 mM Mg^{2+}. A hidden Markov model is used to represent structural transitions between each state

1. P1–P3 strand (transcribed using T7 RNA polymerase): GCGUUCUUAUCAAGAGAAGCACAGGGACUGGC CCGACGAAGCUUCAGCAACCUGCCAAGC GUUCUCUCGC.
2. P3 strand (synthetic strand from IDT, 5NU represents the amino linker): Biotin-GCGAGAGAACGC(5NU)UG.
3. P3–P4 strand (transcribed using T7 RNA polymerase): GCGGGUGCUAAAUCCAGCAAGCUGCGC.
4. P4-P1 strand (synthetic strand from IDT, 5NU represents the amino linker): GCGCAGCUUGGAAGAUAAGAACGC(5NU).

2.2 Single-Molecule FRET

In this section, it is assumed that total internal reflection (TIR) equipment is used for sm-FRET measurements.

1. T50 buffer: 50 mM Tris–HCl, pH 8.1 and 25 mM NaCl.
2. Biotinylated and non-biotinylated bovine serum albumin (BSA) (10 mg/mL) (*see* **Note 1**).
3. Streptavidin: 5 mg/mL to be prepared in the T50 buffer (*see* **Note 2**).
4. Imaging buffer: 50 mM Tris–HCl, pH 8.1, 2.5 mM protocatechuic acid (PCA), 250 nM protocatechuate-3,4-dioxygenase (PCD), and 1 mM 6-hydroxy-2,7,5,8-tetramethylchroman-2-carboxylic acid (TROLOX).
5. Quartz microscope slides of $51 \times 75 \times 1$ mm for prism-type TIR sm-FRET.
6. Fluorescence beads (IFluoSpheres carboxylated 0.2 μm, Invitrogen) for instrument alignment and generation of mapping algorithm to correlate individual donor and acceptor spots.

3 Methods

3.1 Single-Molecule FRET

3.1.1 Quartz Slide Preparation

Clean slides are primordial to obtain meaningful and reliable sm-FRET data. The following protocol can be used independently if quartz slides and glass cover slips are new or recycled.

1. Sonicate quartz slides and cover slips using the following procedure:
   (a) 20 % detergent solution (Hellmanex, Sigma) for 15 min
   (b) MilliQ H_2O for 5 min
   (c) Acetone for 15 min
   (d) MilliQ H_2O for 5 min

(e) 1 M KOH for 15 min

(f) Methanol for 15 min

(g) 1 M KOH for 15 min

(h) MilliQ H_2O for 15 min

2. Dry with N_2 or compressed air.
3. Use a torch flame to remove impurities and moisture from quartz slides and cover slips.
4. Using double-sticky tape, form a channel on quartz slides and place cover slips accordingly on top to create a "sandwich" (*see* **Note 3**).
5. Add H_2O to the channel and look in the microscope to see if the mounted channel is clean before adding immobilization agents. If not, discard the mounted channel and try again with a fresh slide. In the unlikely event that the problem is not solved, repeat the sonication step described above (*see* **Note 4**).

3.1.2 Immobilization of RNA Molecules

1. Dilute tenfold the biotinylated BSA stock in T50 buffer.
2. Add 50 μL of the mixture to the mounted channel.
3. Incubate for 10 min at room temperature to allow BSA binding on the glass surface.
4. To remove unbound material, wash using 60 μL of T50 buffer.
5. Dilute 25-fold the streptavidin solution in T50 buffer. Add 50 μL of this solution to the mounted slide.
6. To remove unbound material, wash using 60 μL of T50 buffer.
7. Add 60 μL of 100 pM solution of fluorescent construct resuspended in T50 buffer.
8. Incubate for 10 min at room temperature to allow RNA binding on coated surface.

3.1.3 Sm-FRET Data Collection and Analysis

To ascertain that single molecules are detected, a mapping algorithm must be used that correlates donor spots with their acceptor counterparts. The use of 200 nm fluorescence beads is helpful to perform such a mapping.

1. Fluorescence data at donor and acceptor wavelengths are collected from single molecules by using total internal reflection microscopy using a 532 nm laser excitation.
2. Data are recorded using an in-house Visual C++ program using integration times ranging from 16 to 100 milliseconds

depending of the sample dynamics. Measurements are usually performed at room temperature.

3. Using the P1–P3 FRET construct, a detailed analysis of time records obtained in the presence of 50 mM EDTA, 0.75 mM Mg^{2+}, and 3 mM Mg^{2+} has shown that the aptamer folds through at least three distinct structural states (*see* Fig. 2b). In the presence of EDTA, a low FRET state is observed corresponding to the unfolded conformer (U). However, the addition of 0.75 magnesium ions allows the aptamer to transit between mid-FRET (F_{Mg}) and high-FRET (F_{NS}) states. A further increase in Mg^{2+} concentration promotes structural transitions mostly between F_{Mg} and F_{NS} states. The addition of SAM further stabilizes the F_{NS} conformer, consistent with the formation of the native state [6]. Biochemical experiments have shown that the $F_{Mg} <-> F_{NS}$ transition detected in the P1–P3 FRET construct corresponds to the helical rotation of the P1 stem [6]. The accumulation of several time traces is required to build population histograms [4].

4 Notes

1. The presence of non-biotinylated BSA reduces nonspecific binding of RNA molecules to plastic containers.
2. This solution should be stored at 4 °C immediately after preparation and use.
3. The channel should have a gap of ~5 mm between both sticky tapes.
4. The use of ultrapure water for single-molecule applications is absolutely crucial since a very low concentration of contaminant could have a drastic effect on the imaging procedure. Thus, the highest quality for water is very important to obtain reproducible data. We have found that commercially available pure water usually gives high-quality results.

Acknowledgements

We thank members of the Penedo and Lafontaine laboratories for critical reading of the manuscript and the National Sciences and Engineering Research Council of Canada (NSERC) for financial support. D.A.L. is a Fonds de Recherche Santé Québec Senior Scholar.

References

1. Breaker RR (2012) Riboswitches and the RNA world. Cold Spring Harb Perspect Biol, 4, 10.1101/cshperspect.a003566
2. Serganov A, Nudler E (2013) A decade of riboswitches. Cell 152:17–24
3. Clegg RM (1992) Fluorescence resonance energy transfer and nucleic acids. Methods Enzymol 211:353–388
4. Cornish PV, Ha T (2007) A survey of single-molecule techniques in chemical biology. ACS Chem Biol 2:53–61
5. Lemay JF, Penedo JC, Tremblay R et al (2006) Folding of the adenine riboswitch. Chem Biol 13:857–868
6. Heppell B, Blouin S, Dussault AM et al (2011) Molecular insights into the ligand-controlled organization of the SAM-I riboswitch. Nat Chem Biol 7:384–392
7. Fiegland LR, Garst AD, Batey RT et al (2012) Single-molecule studies of the lysine riboswitch reveal effector-dependent conformational dynamics of the aptamer domain. Biochemistry 51:9223–9233
8. Haller A, Rieder U, Aigner M et al (2011) Conformational capture of the SAM-II riboswitch. Nat Chem Biol 7:393–400
9. Brenner MD, Scanlan MS, Nahas MK et al (2010) Multivector fluorescence analysis of the xpt guanine riboswitch aptamer domain and the conformational role of guanine. Biochemistry 49:1596–1605
10. Souliere MF, Altman RB, Schwarz V et al (2013) Tuning a riboswitch response through structural extension of a pseudoknot. Proc Natl Acad Sci U S A 110: E3256–E3264
11. Heppell B, Mulhbacher J, Penedo JC et al (2009) Application of fluorescent measurements for characterization of riboswitch-ligand interactions. Methods Mol Biol 540: 25–37
12. Lemay JF, Penedo JC, Mulhbacher J et al (2009) Molecular basis of RNA-mediated gene regulation on the adenine riboswitch by single-molecule approaches. Methods Mol Biol 540:65–76

Chapter 7

Probing of Nascent Riboswitch Transcripts

Adrien Chauvier and Daniel A. Lafontaine

Abstract

The study of biologically significant and native structures is vital to characterize RNA-based regulatory mechanisms. Riboswitches are *cis*-acting RNA molecules that are involved in the biosynthesis and transport of cellular metabolites. Because riboswitches regulate gene expression by modulating their structure, it is vital to employ native probing assays to determine how native riboswitch structures perform highly efficient and specific ligand recognition. By employing RNase H probing, it is possible to determine the accessibility of specific RNA domains in various structural contexts. Herein, we describe how to employ RNase H probing to characterize nascent mRNA riboswitch molecules as a way to obtain information regarding the riboswitch regulation control mechanism.

Key words RNA structure, Nascent mRNA, Native structures, RNase H probing

1 Introduction

Riboswitches are RNA molecules that are involved in the control of transport and/or biosynthesis of cellular metabolites [1]. They control various genetic levels such as transcription, translation, mRNA decay, and splicing [1, 2]. Riboswitches are composed of an aptamer domain and an expression platform that perform ligand recognition and genetic regulation, respectively [1, 2]. Upon metabolite binding, riboswitches experience a variety of structural changes important for the modulation of gene expression. Recently, it was shown that ligand binding to the *pbuE* adenine riboswitch exclusively occurred cotranscriptionally [3–5], suggesting that the intrinsic polarity of transcription is of primordial importance for the ligand-dependent riboswitch genetic regulation. Transcription was also previously shown to be important for the folding of *Tetrahymena* group I intron [6] and RNase P [7], indicating that co-transcriptional folding can be highly important for the outcome of genetic expression in other systems. Given that the biological relevance of co-transcriptional folding may only be emphasized within systems allowing in vitro reconstitution assays, it is highly

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_7,

probable that the general importance of co-transcriptional folding is underestimated by a large degree.

Most probing assays interrogating the structure of RNA molecules (or RNA sub-domains) rely on empirically established renaturation protocols involving variations in temperature and/or salts. By undergoing these various controlled steps, it is expected that RNA molecules are adopting structures that are favored in such defined conditions. However, such condition-enriched RNA structures may not always represent the native structure required for biological activity. Thus, it is highly desirable to probe nascent RNA molecules to ensure that studied conformations are as close as possible to biologically relevant structures found in vivo. However, not all probing assays can be employed given that transcriptional complexes are only active in specific conditions.

Ribonuclease H is a well-known nonspecific endonuclease that catalyzes RNA via a hydrolytic mechanism [8]. In this assay, a DNA oligonucleotide is used to probe the relative accessibility of a region in a given RNA molecule. While a single-stranded region can be targeted by RNase H cleavage, a double-stranded or highly structured region does not allow DNA binding and is thus protected from RNase cleavage. Because RNase H is active in a wide range of experimental conditions, it can be used to probe RNA structural changes in various experimental conditions. For example, RNase H was recently used to characterize the folding pathway of nascent B12 riboswitch where it was determined that co-transcriptional folding is important for riboswitch activity [9]. In this chapter, we describe how RNase H can be used to characterize the structure of nascent riboswitch molecules and how this can be employed to deduce biologically relevant information.

2 Materials

2.1 DNA Templates Required for In Vitro Transcription

1. The DNA template was made by PCR amplification on the genomic DNA from *Escherichia coli* strain MG1655-K12 and consisting of a lacUV5 promoter fused to a riboswitch sequence (*thiM* riboswitch, *see* Fig. 1) followed by a portion of the open reading frame (11 codons). According to the model, the *thiM* riboswitch allows ribosome translation in the absence of ligand. However, ligand binding causes a structural reorganization that sequesters the AUG start codon, thus inhibiting riboswitch translation initiation. The riboswitch DNA template sequence is the following (sequence written 5′–3′ polarity):

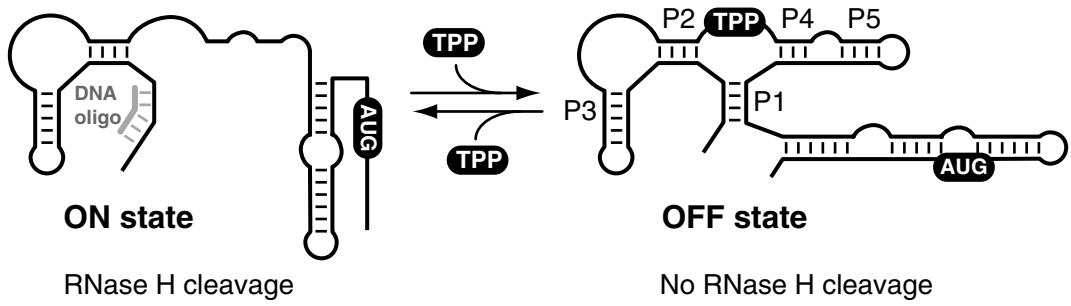


Fig. 1 Schematic representation of the *E. coli thiM* riboswitch. In this model, the riboswitch ON state is adopted in the absence of TPP and allows ribosome translation due to the relative accessibility of the AUG start codon. However, in the presence of TPP, the riboswitch is reorganized and sequesters the AUG start codon, ultimately leading to translation repression. P1 to P5 helical domains are indicated on the riboswitch. According to this model, a DNA oligonucleotide targeting the P1 stem would only be able to hybridize in the absence of TPP

gggcaccccaggctttacactttatgcttccggctcgtataatgtgtggCTGC
GATTTATCATCGCAACCAAACGACTCGGGGTGCCCT
TCTGCGTGAAGGCTGAGAAATACCCGTATCACCTG
ATCTGGATAATGCCAGCGTAGGGAAGTCACGGA
CCACCAGGTCATTGCTTCTTCACGTTATGGCAGGA
G C A A A C T A T G C A A G T C G A C C T G C T
GGGTTCAGCGCAATCT

The promoter is shown in lowercase and the ATG start codon is underlined.

2.2 In Vitro Transcription of RNA Strands Under Single-Round Conditions

1. 5× transcription buffer (100 mM Tris–HCl pH 8.0, 100 mM NaCl, 100 mM $MgCl_2$, 500 μM EDTA, and 70 mM β-mercaptoethanol) (*see* **Note 1**).
2. 100 μM RNA tetranucleotide (5′-CUGC-3′).
3. 2.5 μM ATP and GTP.
4. 1 mM rNTP.
5. 2 μCi [α-^{32}P] UTP.
6. 300 fmol/μL DNA template.
7. 6 mg/mL heparin, RNA polymerase holoenzyme from *Escherichia coli* (Epicentre).
8. 200 μM thiamin pyrophosphate (TPP).

2.3 RNase H Cleavage Assays

1. 1× cleavage buffer (5 mM Tris–HCl pH 8, 20 mM $MgCl_2$, 100 mM KCl, 50 μM EDTA, and 9 mM β-mercaptoethanol) (*see* **Note 1**).
2. RNase H enzyme (Ambion-10 U/μL).
3. The DNA oligonucleotide (100 μM) should target a region that is informative for the study. In the present case, the oligonucleotide is targeting the aptamer region of the *thiM*

riboswitch that should be protected from RNase H cleavage upon TPP binding (*see* Fig. 1). The sequence of the DNA oligonucleotide is the following (sequence written 5′–3′ polarity): CCGAGTCGTT.

4. 2× stop solution: 95 % formamide, 18 mM EDTA, and 0.02 % SDS.

3 Methods

Because RNases are prevalent and very hard to inactivate, it is thus important to employ RNase free solutions. Moreover, ensure that all reagents have been equilibrated at room temperature before mixing together.

3.1 In Vitro Transcription of RNA Strands Under Single-Round Conditions

1. In a total volume of 7.5 μL, add 1.25 μL of ATP/GTP mixture, 0.85 μL CUGC tetranucleotide, 2 μL 5× transcription buffer, 1 μL DNA template, and 1 μL of 200 μM TPP when necessary.
2. Incubate the sample for 5 min at 37 °C (*see* **Note 2**).
3. In the reaction mixture, add a volume of 2.5 μL containing 10 μCi [α-^{32}P] UTP and 0.2 U of *E. coli* RNA polymerase (homemade or commercially available).
4. Incubate the sample for 15 min at 37 °C.
5. In the reaction mixture, add a 10 μL solution containing 2 μL transcription buffer, 1 μL rNTPs, 1.5 μL heparin, and 5.5 μL water (*see* **Note 3**).
6. Incubate the sample for 15 min at 37 °C.
7. At this point, the presence of heparin is ensuring that transcription is stopped after complete RNA strand synthesis. It is important to avoid adding denaturing agents since this would disrupt native RNA interactions and therefore mislead conclusions.

3.2 RNase H Cleavage Assays

1. Dilute RNase H just prior to use in ice-cold cleavage buffer to 0.12 U/μL (*see* **Note 4**).
2. In a fresh tube, incubate 8 μL of the transcription reaction with 2 μL of DNA oligonucleotide.
3. Incubate the reaction mixture for 5 min at 37 °C.
4. Add 10 μL of RNase H dilution and incubate for 5 min at 37 °C.
5. Quench reaction with 20 μL of 2× stop solution.
6. Load on a denaturing gel to separate uncleaved from cleaved riboswitch transcripts.
7. The result of a typical reaction is shown in Fig. 2. In this experiment, it can be observed that RNase H cleavage is only observed

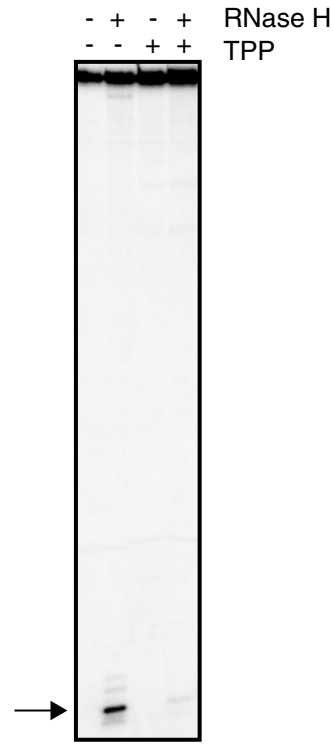


Fig. 2 RNase H cleavage of the *thiM* riboswitch as a function of TPP binding. The *arrow* indicates the production of the 5′ product due to RNase H cleavage activity in the absence of TPP. The cleavage is highly reduced in the presence of TPP, indicating that the DNA oligonucleotide does not bind to the riboswitch

in the absence of TPP, as expected from the riboswitch model in Fig. 1. The absence of cleavage in the presence of TPP indicates that TPP binding to the aptamer domain resulted in a stable structure that does not allow DNA binding and thus RNase H cleavage.

4 Notes

1. The β-mercaptoethanol has to be added just before setting up the transcription reaction.
2. Preincubate the mixture at 37 °C to allow temperature equilibration.
3. Prepare the mixture and equilibrate at 37 °C before using.
4. As the RNase H enzyme is sensitive to freeze-thaw cycles, it is best to prepare a dilution and to equilibrate it at room temperature before using it.

Acknowledgements

We thank members of the Lafontaine laboratory for critical reading of the manuscript and the National Sciences and Engineering Research Council of Canada (NSERC) for financial support. D.A.L. is a Fonds de Recherche Santé Québec Senior Scholar.

References

1. Breaker RR (2012) Riboswitches and the RNA world. Cold Spring Harb Perspect Biol 4:a003566. doi:10.1101/cshperspect.a003566
2. Serganov A, Nudler E (2013) A decade of riboswitches. Cell 152:17–24
3. Lemay JF, Penedo JC, Tremblay R et al (2006) Folding of the adenine riboswitch. Chem Biol 13:857–868
4. Lemay JF, Desnoyers G, Blouin S et al (2011) Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. PLoS Genet 7:e1001278
5. Rieder R, Lang K, Graber D et al (2007) Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. Chembiochem 8:896–902
6. Koduvayur SP, Woodson SA (2004) Intracellular folding of the tetrahymena group I intron depends on exon sequence and promoter choice. RNA 10:1526–1532
7. Pan T, Artsimovitch I, Fang XW et al (1999) Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. Proc Natl Acad Sci U S A 96:9545–9550
8. Zarrinkar PP, Williamson JR (1994) Kinetic intermediates in RNA folding. Science 265: 918–924
9. Perdrizet GA II, Artsimovitch I, Furman R et al (2012) Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. Proc Natl Acad Sci U S A 109:3323–3328

Chapter 8

Functional Studies of DNA-Protein Interactions Using FRET Techniques

Simon Blouin, Timothy D. Craggs, Daniel A. Lafontaine, and J. Carlos Penedo

Abstract

Protein-DNA interactions underpin life and play key roles in all cellular processes and functions including DNA transcription, packaging, replication, and repair. Identifying and examining the nature of these interactions is therefore a crucial prerequisite to understand the molecular basis of how these fundamental processes take place. The application of fluorescence techniques and in particular fluorescence resonance energy transfer (FRET) to provide structural and kinetic information has experienced a stunning growth during the past decade. This has been mostly promoted by new advances in the preparation of dye-labeled nucleic acids and proteins and in optical sensitivity, where its implementation at the level of individual molecules has opened a new biophysical frontier. Nowadays, the application of FRET-based techniques to the analysis of protein-DNA interactions spans from the classical steady-state and time-resolved methods averaging over large ensembles to the analysis of distances, conformational changes, and enzymatic reactions in individual protein-DNA complexes. This chapter introduces the practical aspects of applying these methods for the study of protein-DNA interactions.

Key words Protein-DNA interactions, Fluorescence spectroscopy, Förster resonance energy transfer, Time-resolved fluorescence, Single-molecule detection

1 Introduction

Protein-DNA interactions are widespread. Understanding the molecular basis of these crucial biological mechanisms requires therefore a detailed analysis of the nucleoprotein complex structure and the dynamic interactions that govern its assembly and function. Owing to their sensitivity and the recent advances in site-specific dye-labeling methods in both nucleic acids and proteins, fluorescence detection-based biophysical assays have proven to be very powerful and versatile techniques to probe the dynamics and function of protein-DNA complexes [1–3]. The appeal of these fluorescence-based approaches relies on the extreme sensitivity of a fluorescence probe to its environment [4], the possibility of monitoring the fluorescence signal

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_8,

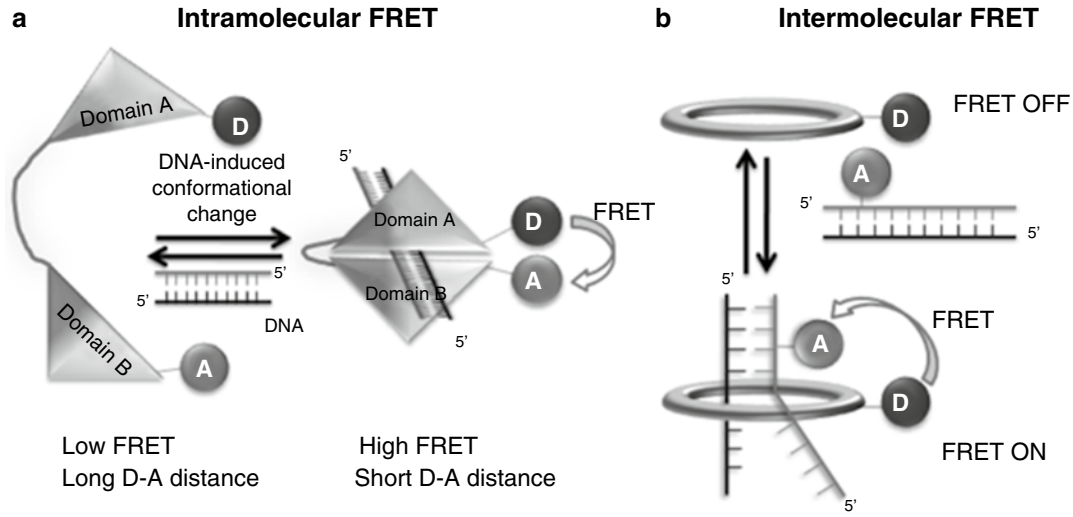


Fig. 1 Protein-DNA interactions can be monitored by intramolecular and intermolecular FRET assays. (**a**) Intramolecular FRET assays are based on the functionalization of the same biomolecule with both donor and acceptor fluorescence labels. Interaction with the partner biomolecule induces a conformational change in the host molecule (i.e., between an open and a closed state) that modifies the donor–acceptor distance and therefore the FRET value. (**b**) Intermolecular FRET assays are engineered with each fluorescence label in different biomolecules. When the two interacting molecules fuse together and form a stable complex, FRET occurs. On the other hand, when the complex dissociate, the FRET value diminishes

continuously in real time to provide accurate kinetic data, and, when combined with Förster resonance energy transfer (FRET), the ability to provide insights into the structural basis of DNA or protein-induced conformational rearrangements [5–7].

The basic idea underlying any FRET experiment designed to study a particular protein-DNA complex relies on the site-specific labeling with a donor and an acceptor dye, located both in the same biomolecule (intramolecular FRET, Fig. 1a) or one FRET dye in each interacting partner (intermolecular FRET, Fig. 1b). Direct optical excitation of the donor dye results in fast energy transfer to the FRET acceptor, which emits fluorescence at a longer wavelength. The efficiency of this process depends on the sixth power of the average distance between the donor and the acceptor dye [8–10] and thus the changes in fluorescence intensity from donor and acceptor can be used to monitor the interaction between proteins and their DNA substrate with extreme sensitivity and accuracy [11–13].

Intramolecular FRET assays where both dyes are located in the same biomolecule have been extensively used to monitor protein-induced conformational changes in the DNA substrate and to determine the global structure and assembly dynamics of a variety of nucleoprotein complexes. These studies include analysis of DNA bending upon interaction with a range of DNA-binding proteins such as the TATA protein [14], the high-mobility group box HMG [5, 15], the integration host factor protein IHF [16, 17], the catabolite activator protein CAP [7, 13], and the 5′ Flap

Endonuclease FEN1 [18]. DNA strand exchange proteins such as RecA [19, 20] and its eukaryotic homologs hRad51 and scRad51 [21], single-stranded binding proteins [22], RecBCD-like nucleases [23], and Holliday junction-resolving enzymes such as archaeal Hjc [24] have been the subject of intensive research using FRET-related techniques both at ensemble and single-molecule level. FRET has also been used to investigate the relative orientation of single-stranded DNA template primers with respect to the Klenow fragment of *E. coli* DNA polymerase I [25]. More recently, FRET methods have been specifically developed to monitor the movement of RNA polymerase (RNAP) relative to DNA during transcription [26–28] and define the three-dimensional structure of transcription complexes in solution [29]. Apart from those studies focused on the understanding of the molecular basis of DNA recognition outlined above, the investigation of cleavage reactions of nucleic acids catalyzed by a variety of enzymes is another major area where FRET techniques are already providing a wealth of kinetic information [30]. Usually, the efficiency of the enzymatic cleavage process is determined using a DNA substrate doubly labeled with donor and acceptor fluorophores. In the absence of enzymatic reaction, the proximity of both dyes enables efficient energy transfer from the donor to the acceptor, thereby decreasing the fluorescence intensity of the donor moiety [31]. Upon incubation with the enzyme, cleavage of the DNA substrate leads to the separation of the donor and acceptor dyes, with the concomitant cease of energy transfer and increase of the donor fluorescence. The main advantage of this fluorimetric assay when compared to more conventional biochemical techniques such as radiolabeling-based electrophoresis or ELISA-based techniques relies on its continuous character, so that the cleavage reaction can be monitored from the initial steps in real-time with no need for extensive sample handling. Following this approach, the kinetics of restriction endonucleases such as *PaeR7* [32], *Eco*RV [33], S1 nuclease [34], and Endonuclease V [35] have all been quantified by FRET techniques. Here, we present methods and protocols for FRET experiments that permit functional studies of protein-DNA interactions including biomolecular conformational changes and cleavage assays.

1.1 Fluorescence Resonance Energy Transfer (FRET)

FRET is a nonradiative process whereby an excited donor fluorophore D transfers energy to a ground state acceptor A (Fig. 2a) as a result of a through-space coupling of their transition dipoles [8–10]. According to Förster's theory, the energy transfer efficiency E depends on the inverse sixth power of the distance, R, separating the D–A pair:

$$E = \frac{R_o^6}{R_o^6 + R^6} = \frac{1}{1 + \left(\frac{R}{R_o}\right)^6} \quad (1)$$

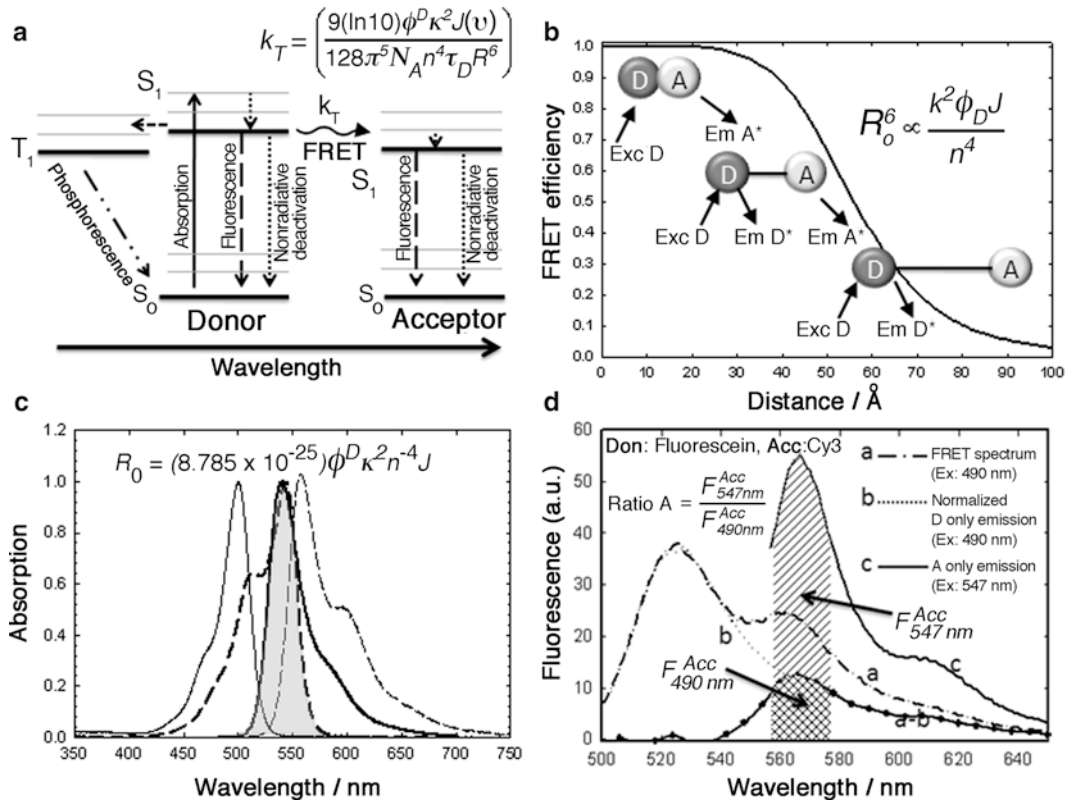


Fig. 2 FRET principle: (**a**) Absorption of a photon by the donor excites an electron to an upper state S_1. The excited donor can spontaneously lose its excess energy and decay to the ground state by a combination of competing processes including fluorescence emission, nonradiative deactivation by interaction with solvent molecules, intersystem crossing to the triplet state T_1 or, in the presence of a nearby acceptor molecule, by resonance energy transfer. The probability of energy transfer, proportional to k_T, decreases rapidly when the distance increases (*inset equation*). (**b**) Distance-dependence of the energy transfer efficiency according to Förster theory (Eq. 1) for the fluorescein (donor)-Cy3 (acceptor) FRET pair. The Förster radius R_0, (56 Å) that represents the distance at which 50 % of the donor deactivates by energy transfer, is calculated according to Eq. 2. Because of the sixth-power dependence of the energy transfer with the distance, FRET is more sensitive for changes in distance close to the R_0 value. (**c**) Absorption and fluorescence spectra for the FRET pair fluorescein-Cy3. The overlapping region between donor emission and acceptor absorption is illustrated by the *shaded region* (*inset*: expression for R_0 in terms of wavenumbers). (**d**) Example of steady-state FRET data analysis, for the particular case of fluorescein-Cy3 FRET pair, following the enhanced acceptor emission method. The fluorescence emission of the donor in a donor-only labelled biomolecule (b, *dotted line*) is obtained with excitation at its absorption maximum and normalized in the region 510–530 nm to the experimental FRET spectrum (a, *dashed dotted line*) excited at the same wavelength. The "pure acceptor emission" (a – b) is obtained by subtracting the normalized donor emission from the experimental FRET spectrum. A fluorescence spectrum of the acceptor only (c, *dashed line*) is also taken with excitation at is absorption maximum (547 nm), where no donor absorption takes place. The *ratioA* parameter, proportional to the FRET efficiency, is calculated from the ratio $F_{547}^{acc}/F_{490}^{acc}$

where R_0 is the distance between donor and acceptor at which 50 % of the excited D* molecules decay by energy transfer (Fig. 2b), while the other half decays through other radiative and nonradiative deactivation channels. This critical transfer distance, the so-called

Förster distance, is a characteristic property of the donor–acceptor pair used and can be calculated for a particular FRET pair from the spectral and photophysical properties of the donor and acceptor partners following the expression:

$$R_o^6 = \frac{9000 \ln 10 k^2 Q_D J}{128 \pi^5 n^4 N_A} \quad (2)$$

where Q_D is the donor's fluorescence quantum yield, n is the refractive index, N_A is Avogadro's number, k^2 is the orientation factor, and J represents the spectral overlap between the donor fluorescence emission spectrum and the acceptor absorption spectrum (Fig. 2c) obtained from the expression

$$J = \int F_D(\lambda) \varepsilon_A(\lambda) \lambda^4 d\lambda \quad (3)$$

where $F_D(\lambda)$ is the fluorescence emission of the donor normalized to a total peak area of 1, and $\varepsilon_A(\lambda)$ is the absorption spectrum of the acceptor normalized to the molar extinction coefficient. The quantum yield of the donor chromophore needs to be determined in the context of the biomolecule as this can markedly differ from that in free solution. The refractive index of the medium is assigned a value of 1.4 for biomolecules in aqueous solution. The k^2 factor describing the orientation of the donor–acceptor transition dipole can range from 0 for perpendicular orientation to 4 for parallel. When complete averaging of the relative orientation of the dyes is achieved during the excited state lifetime of the donor, k^2 adopts a value of two-thirds, which has proven to be a reasonable approximation in most bimolecular environments. For a more detail description on the influence of k^2 on the measured distance values, which is beyond the scope of this chapter, we refer to a recent review by van der Meer [36]. As an example, the distance dependence of the FRET efficiency for the fluorescein-Cy3 donor acceptor pair is illustrated in Fig. 2b (note that for $R = R_o$, the FRET efficiency is 0.5).

Experimentally, the most direct approach to measure FRET efficiencies is based on steady-state fluorescence techniques by measuring the intensity of the donor in the absence I_D and in the presence of acceptor I_{DA} according to Eq. 4. From here it is possible to recover the donor–acceptor distance following Eq. 5:

$$E = 1 - \frac{I_{DA}}{I_D} \quad (4)$$

$$R_{DA} = R_o \left(\frac{I_{DA}}{I_D - I_{DA}} \right)^{1/6} \quad (5)$$

1.2 Time-Resolved Fluorescence Resonance Energy Transfer

It is very often the case that interacting biomolecules are flexible structures showing a broad range of donor–acceptor distances. In these cases, the efficiency of energy transfer calculated from steady-state measurements overestimates the correct distances and

represents an average over all possible conformational mixtures present in solution. When the number of different conformations, the relative equilibrium populations of each conformation and their corresponding true distances need to be resolved, time-resolved FRET (tr-FRET) techniques are the best choice [37]. In a time-resolved FRET experiment, the measured parameter is the fluorescence lifetime of the donor, which for most organic dyes ranges in the picosecond to nanosecond time range. In the absence and presence of an acceptor, the fluorescence intensity will decay according to the expressions in Eqs. 6 and 7, respectively:

$$I_{\mathrm{D}}(t) = I_{\mathrm{o}} \exp\left(\frac{-t}{\tau_{\mathrm{D}}}\right) \tag{6}$$

$$I_{\mathrm{DA}}(t) = I_{\mathrm{o}} \exp\left[\left(\frac{-t}{\tau_{\mathrm{D}}}\right)\left(1 + \left(\frac{R_{\mathrm{o}}}{R}\right)^{6}\right)\right] \tag{7}$$

where τ_{D} represents the lifetime of donor in the absence of acceptor and the second component in the exponential term represents the energy transfer rate. However, it is normally the case that a fluorescence probe attached to a biomolecule can adopt multiple conformations, each of them with an intrinsic fluorescence lifetime. In this case, the fluorescence decay should be modeled as a distribution of lifetimes as represented in Eq. 8, where *fi* is the relative population of each species and *Pi*(*r*) represents the distance distribution:

$$I_{\mathrm{DA}}(t) = I_{\mathrm{o}} \sum_{i} f_i \int P_i(R) \exp\left[-\frac{t}{\tau_{\mathrm{D}}}\left(1 + \left(\frac{R_{\mathrm{o}}}{R}\right)^{6}\right)\right] \mathrm{d}R \tag{8}$$

The two existing methodologies for measuring time-resolved FRET, phase-modulation and time-correlated single-photon counting (TCSPC), together with the procedures for data analysis, have been recently reviewed by Klostermeier and Millar [37]. The expressions outlined above are commonly applied to study conformational changes and extract distance information in proteins or DNA. However, in enzymatic cleavage assays followed by tr-FRET, it is usually more practical to analyze the changes in the average donor lifetime (Eq. 9) as a function of the progress of the cleavage reaction (Fig. 3d). In these assays, a DNA substrate labeled with donor and acceptor is engineered (Fig. 3a) so that, due to the enzymatic reaction, the acceptor moiety will be released with the concomitant increase in donor lifetime and breakdown of FRET over time:

$$\tau = \frac{\sum_{i} \alpha_i \tau_i^2}{\sum_{i} \alpha_i \tau_i} \tag{9}$$

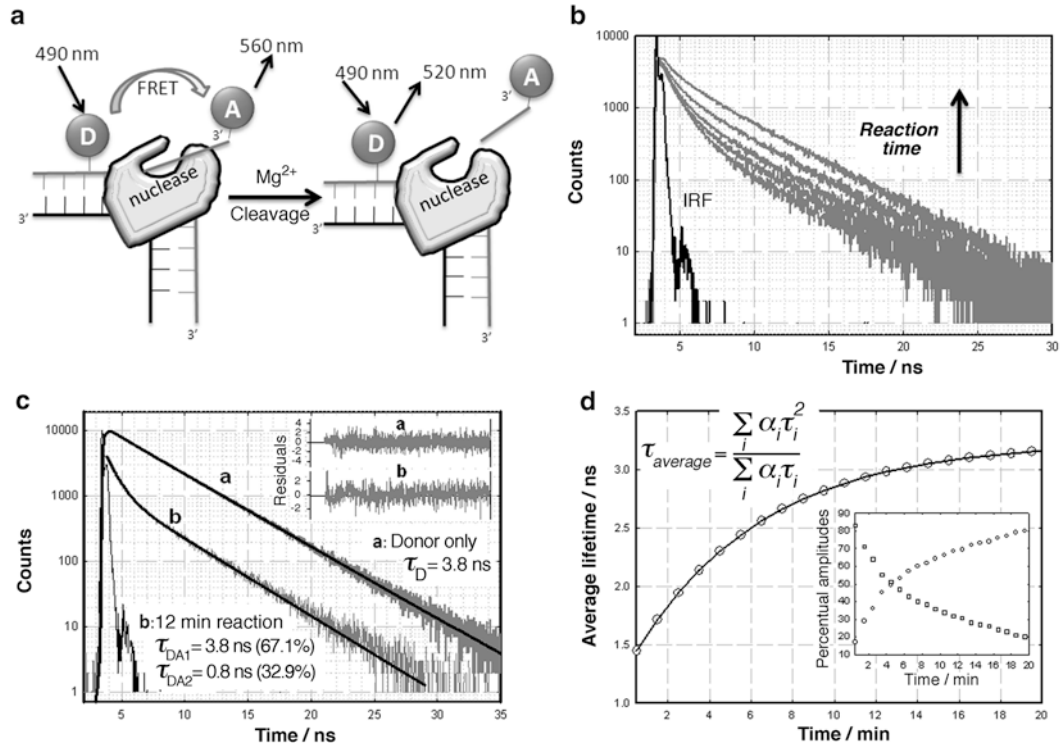


Fig. 3 Example of an enzymatic DNA cleavage assay monitored by time-resolved fluorescence resonance energy transfer. (**a**) A 3′-flap duplex DNA is labelled with donor (fluorescein) and acceptor (Cy3) groups. In the presence of magnesium divalent ions, binding of structure-specific endonucleases such as XPF (xeroderma pigmentosum, complementation group F) induces cleavage of the 3′-flap with the subsequent release of the FRET acceptor and FRET breakdown (Penedo et al., unpublished data). (**b**) Monitoring the nuclease-induced cleavage of the DNA construct shown in (**a**) by time-resolved fluorescence spectroscopy. As the cleavage progresses, the biexponential decay together with time-dependent contributions from both product (donor only, $\tau_D = 3.8$ ns) and uncleaved substrate (donor + acceptor, $\tau_{FRET} = 0.8$ ns) evolves to a monoexponential donor-only decay. Each fluorescence decay trace was taken at fixed intervals of time (0.5 s) during the progression of the cleavage reaction. The excitation was performed by a 475 nm picosecond pulsed-diode laser (Edinburgh Instruments Ltd, UK) and the emission monocromator was placed at 520 nm, the donor emission maximum. Each fluorescence trace represents the time-correlated single photon-counting histogram accumulated during 0.5 s. The instrument-response function (IRF) used to deconvolute the experimental decays is also shown. (**c**) Comparison between the lifetime decays of the donor only control and the doubly labeled DNA construct after 12-min progress of the cleavage reaction. The lifetime traces have been fitted to a monoexponential decay in the case of the donor only DNA control and a biexponential decay following the expression: $F(t) = F_0(t) + A_1\exp(-t/\tau_1) + A_2\exp(-t/\tau_2)$ for the doubly labeled substrate. The pre-exponential factors A_1 and A_2, representing the amplitudes of each exponential at time zero, are proportional to the amounts of product and uncleaved substrate. As an example at 12-min reaction time, the remaining uncut substrate constitutes ~33 % of the total intensity. *Inset*: Residuals obtained from the fitting of the experimental decays to a monoexponential function (*a*, donor-only control) and a biexponential function (*b*, doubly labeled substrate). (**d**) Plot showing the average lifetime calculated according to Eq. 9 as a function of cleavage time. *Inset*: Percentual amplitudes corresponding to product and uncut substrate as a function of cleavage time. The complete dataset of decay traces obtained every 0.5 s was fitted to biexponential decay following global analysis (Edinburgh Instruments Ltd, UK)

1.3 Single-Molecule FRET

More recently, FRET experiments at single-molecule level (Sm-FRET) have become possible providing information on protein-nucleic acid dynamics that was previously hidden when using bulk-solution methods [38–41]. Single-molecule FRET techniques provide a completely new approach to study the structure-dynamics-function relationship in protein-nucleic acid complexes, allowing the identification of subpopulations in a heterogeneous mixture, the analysis of protein-DNA complexes lifetimes, and the recovery of FRET efficiency distributions, with the advantage compared to time-resolved FRET that it does not require any assumption about the shape of the distance distribution. The efficiency of energy transfer in a single-molecule experiment is usually calculated from the expression

$$E = \frac{I_A}{I_D + \alpha I_A} \tag{10}$$

where I_D and I_A represent the fluorescence intensity of the donor and acceptor, respectively. The α parameter corrects for the donor leakage into the acceptor channel and direct excitation of the acceptor at the donor excitation wavelength. All these corrections can be done with donor- and acceptor-only labeled species [39, 42]. Single-molecule FRET experiments can be performed either on freely diffusing molecules [43, 44], or on immobilized molecules to record the trajectory of the molecule for extended periods of time [42, 45], but in this case care must be taken to ensure that the immobilization techniques are compatible with the biomolecule(s) under study [46]. Single-molecule FRET studies on protein-DNA interactions have mostly focused on the unwinding of DNA by helicases [47, 48], DNA damage [49, 50], and the analysis of the relative movement of RNAP on the DNA template during the initiation and elongation steps of the transcription processes [28, 29].

2 Materials

2.1 Instrumentation

For steady-state FRET experiments, any scientific-grade commercial fluorimeter equipped with conventional Xe lamps and with real-time correction for fluctuations in lamp intensity will be appropriate. For time-resolved FRET two different experimental approaches can be used [37], either the frequency domain method or the time domain method. In the frequency domain method, the sample is excited with a sinusoidally modulated laser of certain frequency and the fluorescence lifetime is extracted from the phase delay and the demodulation of the emitted light. In the time domain method, the sample is excited with a short laser pulse, typically of duration ~80 ps or less, and the fluorescence emission is detected with a fast photomultiplier or multichannel plate (MCP)

and a time-correlated single-photon counting system (TCSPC). Either in steady-state or in time-resolved conditions, the fluorescence signal must be collected at magic angle conditions, with the emission polarizer placed at 54.7° relative to the vertically polarized excitation, to avoid artifacts due to molecular reorientation.

For single-molecule experiments the two most common approaches are fluorescence correlation spectroscopy (FCS), where molecules are allowed to freely diffuse through the excitation volume, and those techniques requiring the immobilization of the biomolecule under study to allow observation for longer periods of time: total-internal reflection (Fig. 4) and scanning confocal microscopy. Single-molecule instrumentation for FCS is already commercially available from a range of manufacturers (Leica, Olympus, Picoquant) and the details of single-molecule FCS instrumentation can be found in several recent reviews [43, 44]. For immobilization single-molecule experiments, there is no specific commercial equipment as for FCS, so scientists working in the field need to build their own equipment. It will be assumed in the following that a single-molecule total internal reflection (Sm-TIR) setup is already available. Single-molecule instrumentation share in common the need for a laser excitation source, an inverted microscope with oil- or water-immersion objectives depending on the particular setup (*see* **Note 1**), and a detection system. Depending whether the aim is to monitor many molecules simultaneously or just one but with higher time resolution, the acquisition system will be based on two-dimensional detectors, such as intensified or electron-multiplying CCD cameras (Andor, UK, or Princeton Instruments, USA) or point-detection devices such as avalanche photodiodes APDs (Perkin Elmer, USA). The advantage of FCS is that it is a true solution-based technique, and therefore free of perturbations by biomolecule-surface interactions. The main disadvantage is that the observation time is limited to a few milliseconds, the diffusion time of the biomolecule through the confocal volume. To observe single-molecule events for long periods of time, immobilization of the single molecule on an appropriate surface is the method of choice. Several different immobilization methods have been used depending on the type of biomolecule. These include single-point attachment on a glass or quartz surface via BSA-streptavidin interactions [51], poly(ethylene glycol) (PEG)-coated surface [52, 53] (*see* **Note 2**), His-NTA interactions [54, 55], trapping inside porous gel matrices of agarose or polyacrylamide [50], and inside biomimetic membranes such as vesicles [56] and nanopores [47]. Independent of the immobilization method, specific control experiments should be made for each particular system to ensure that the biomolecule is not interacting with the surface, so that any observed heterogeneity is intrinsic and does not arise from molecular interactions with the surface [46].

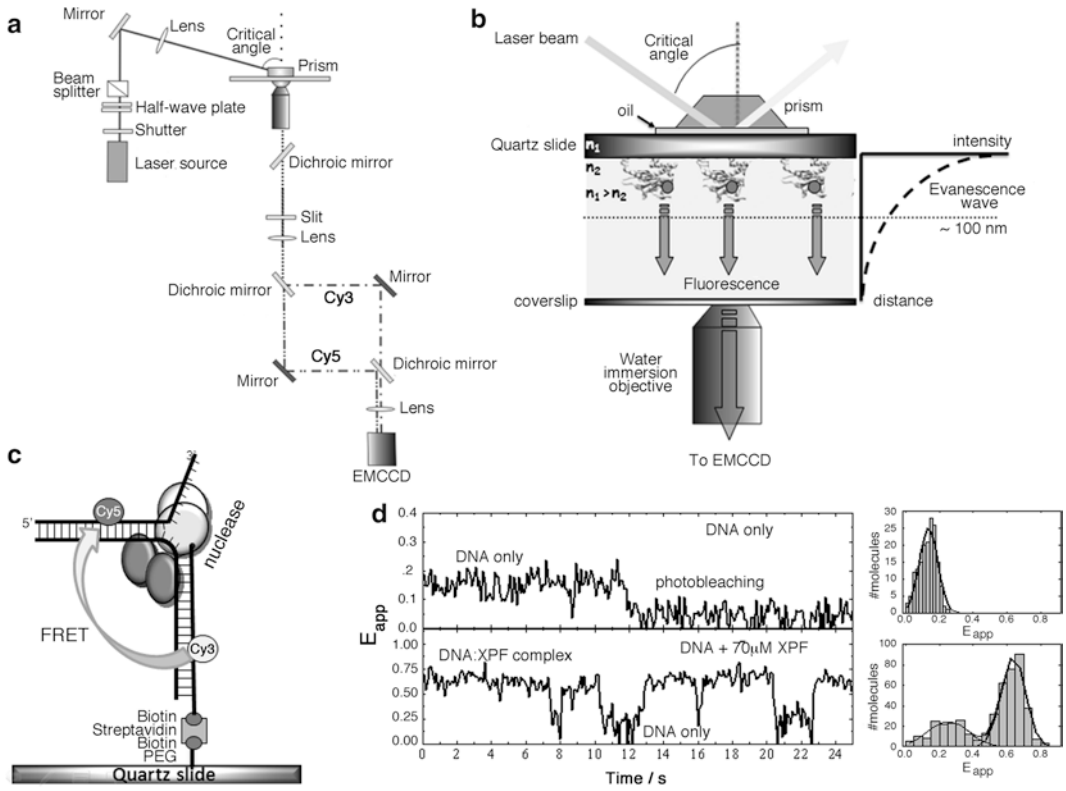


Fig. 4 Single-molecule FRET assays for protein-DNA interactions by total-internal reflection spectroscopy. (**a**) Schematics of a single-molecule TIR microscope for single-molecule FRET. Briefly, the sample is excited by a continuous wave diode laser (532 nm, Crystalaser) at a critical angle of ~65° by using a prism-type setup. The fluorescence obtained from the sample is collected by a water immersion objective (60×, NA 1.2, WD: 0.17–0.25) mounted on an inverted microscope (Olympus IX71). A dichroic mirror located underneath the objective is used to block the laser excitation light and allow the fluorescence from donor (Cy3) and acceptor (Cy5) to reach a 3 mm slit (Thorlabs, UK). Using a combination of reflecting and dichroic mirrors, the fluorescence is decomposed in its donor (reflected) and acceptor (transmitted) components and spatially separated so that the donor reaches the left-hand side of the CCD camera ship (Andor Technology, UK) and the acceptor reaches the right-hand side. (**b**) Principle of total internal reflection illumination. Refractive index differences between the quartz slide (n_1) and water phases (n_2) modulate how much light is refracted or reflected at the interface as a function of angle of incidence of the laser beam. At a specific critical angle, the beam is completely reflected from the quartz/water interface. This reflection generates an evanescence wave (~100 nm) in the aqueous medium, which is used to excite the molecules directly attached to the quartz inner surface. Because the intensity of this evanescence wave decays exponentially with the distance from the surface, very small excitation volumes can be easily achieved, increasing the signal-to-noise ratio to appropriate values for single-molecule detection. (**c**) Doubly labeled DNA carrying a 3′-flap engineered for single-molecule binding studies. The construct is immobilized in the quartz surface, which was previously coated with biotinylated-PEG, via biotin-streptavidin interactions. Binding of the structure-specific endonuclease XPF induces a conformational change in the DNA that reduces the distance donor–acceptor, thus increasing the FRET signal as it can be seen in the single-molecule traces represented in (**d**) in the presence and absence of protein. Fluctuations in the FRET signal (*bottom plot*) are indicative of protein binding-dissociation events. FRET-population histograms for each trace are also shown

2.2 Chemicals

Functional studies of protein-DNA interactions by fluorescence resonance energy transfer require the labeling of the DNA or the protein [57], or both, with a FRET pair suitable for the distance range to be measured, and the subsequent purification steps to get reliable FRET efficiency and distance values. In the particular case of single-molecule applications we should emphasize the need for reagents, buffers, and even water (*see* **Note 3**), absolutely free of fluorescence contaminants. Common buffers such as Tris, phosphates, widely used monovalent and divalent ions solutions, and denaturing agents can be obtained with excellent purity from most suppliers. Specific single-molecule additives such as surface-coating molecules (BSA, PEG), oxygen scavenger proteins (glucose oxidase and catalase) and triplet-state quenchers (2-Mercaptoethanol, TROLOX) can be obtained with high purity from more specialized sources (Roche Diagnostics, Pierce Biotechnology, New England Biolabs).

2.2.1 Synthetic DNA

Synthesis of custom DNA oligonucleotides is a procedure now outsourced by most laboratories. Companies such as Operon, DNA Integrated Technologies and Invitrogen offer custom DNA synthesis at micromole and nanomole scales up to 145–150 nucleotides long. For single-molecule immobilization studies biotin is usually incorporated during chemical synthesis at the 5′ or 3′ end.

2.2.2 Fluorophores

1. A variety of fluorescence molecules can be used as donor and acceptor pairs for FRET studies [57, 58]. Table 1 summarizes the spectroscopic properties of most widely used FRET pairs. Fluorescein-tetramethylrhodamine and fluorescein-Cy3 are very popular FRET pairs for bulk-solution investigations. However, for single-molecule studies, especially under immobilization conditions, Cy3-Cy5 is the most accepted FRET pair because of its wider spectral separation when compared to fluorescein-Cy3 and its higher photostability.
2. The same companies that provide synthetic DNA oligonucleotides can also provide fluorescently labeled-DNA with most common donor and acceptor dyes (Fluorescein, Cyanine, and Alexa dyes) at the 5′ and 3′ end terminals and also internal fluorescein-dT. However, the maximum length available for these fluorescently labeled constructs is ~60–70 nucleotides.

2.2.3 DNA Post-synthesis Fluorescence Labeling

An alternative to the incorporation of fluorescence labels on the DNA during chemical synthesis relies on the post-synthesis coupling of a succinimidyl ester (NHS) dye derivative (GE Healthcare, Molecular Probes, Invitrogen) to a synthetic DNA carrying a primary amino group at a specific position. This method is widely used to incorporate internal fluorescence labels other than the commercially available fluorescein. For this labeling procedure the following reagents are required:

Table 1
Summary of spectroscopic properties, donor lifetime values, and Förster radius R_0 for some of the most common FRET pairs used in conventional steady-state and time-resolved FRET, together with those particularly useful in single-molecule FRET applications because of their enhanced photostability

Donor	λ_{exc} (nm)/λ_{em} (nm)	τ_{Donor}/ns	Acceptor	λ_{exc} (nm)/λ_{em} (nm)	R_0
EDANS	336/468	13.0	DBACYL	471/non-fluorescent	33
IAEDANS	336/490	12.0	Fluorescein	494/521	46
GFP	395,475/509	2.6	DsRed	558/583	47
CFP	439/476	3.4(0.6), 1.3(0.4)[a]	YFP	514/527	50
Fluorescein	494/521	4.1[b]	TMR	550/572	55
Fluorescein	494/521	4.1[b]	Cy3	547/565	56
FRET pairs suitable for single-molecule studies					
Cy3	547/560	0.3	Cy5	647/667	53
Cy5	647/667	1	Cy5.5	675/694	–
Alexa 488	495/519	3.9	Alexa 568	578/603	62
Alexa 546	556/573	4.0	Alexa 594	590/617	71
Alexa 594	590/617	3.9	Alexa 647	650/668	85

[a]The number in brackets represents the normalized pre-exponential factors or amplitudes at time zero obtained from time-resolved experiments
[b]Fluorescein incorporated in a duplex DNA shows a second component with a lifetime of ~0.8 ns and a small amplitude (<15 %) indicating different rearrangements of the fluorophore in the DNA duplex

1. Synthetic single-stranded DNA oligonucleotides carrying a primary amino group modification (Operon, IDT).
2. Cy3-NHS, Cy5-NHS, or other appropriate dye carrying a succinimidyl ester group.
3. Dimethyl sulfoxide (e.g., Fisher Scientific).
4. Labeling buffer: 100 mM sodium tetraborate (Sigma), pH 8.5. Keep this solution aliquoted in the freezer for long-term storage and avoid long-term exposition to air, which could change the buffer pH.
5. Precipitation solution: 3 M sodium acetate and 100 % ethanol. Keep these solutions at room temperature.

2.2.4 Protein Labeling

Currently the most common approach is cysteine labeling with a thiol-reactive maleimide derivative of the dye [57].

1. Protein with a surface-exposed cysteine residue.
2. Labeling buffer: 100 mM phosphate-buffered saline between pH 7.0 and 7.5. Tris–HCl or HEPES buffer could also be used.
3. Dimethyl sulfoxide (e.g., Fisher Scientific).

4. Maleimide derivative of the fluorescence dye (Invitrogen, Molecular Probes, GE Healthcare).
5. Dithiothreitol (DTT) (Sigma) or Tris(2-carboxyethyl)phosphine HCl (TCEP) (Sigma) as reducing agents for disulfide bonds. TCEP has the advantages of providing no pungent odour and it is often not needed to remove TCEP before thiol modification using iodoacetamides or maleimides.
6. Glutathione or mercaptoethanol (Sigma).

2.2.5 Fluorescently Labeled DNA Purification by Polyacrylamide Gel Electrophoresis (PAGE)

1. 5× TBE: 450 mM Tris–HCl, 450 mM borate, and 10 mM ethylenediamine tetraacetic acid disodium salt (EDTA).
2. Dilution buffer 1×: 7 M urea dissolved in 1× TBE (store at room temperature).
3. 20 % acrylamide/bis solution (19:1) with 7 M urea in 1× TBE. This solution is neurotoxic when not polymerized and so care should be taken not to receive exposure. Keep this solution in the dark at room temperature.
4. Ammonium persulfate (APS): Prepare 10 % (w/v) solution in water. Keep this solution in the dark at 4 °C (*see* **Note 4**).
5. *N*,*N*,*N*′,*N*′-Tetramethyl-ethylenediamine (TEMED).
6. Running buffer (1×): 1× TBE.
7. Molecular weight markers: Xylene cyanol FF 0.02 % and bromophenol blue 0.02 % dissolved in pure formamide.

2.2.6 DNA Recovery

1. Electroeluter from Harvard lab shops (http://www.mcb.harvard.edu/bioshop).
2. Ammonium acetate (NH_4OAc) 8 M in water. Keep the solution at room temperature.
3. Electro-eluter running buffer: 0.25× TBE.
4. Precipitation solution: 3 M NaOAc and 100 % ethanol. Keep these solutions at room temperature.
5. Elution buffer for crush and soak elution: 0.5 M NH_4OAc, 1 mM EDTA, and 0.1 % (w/v) sodium dodecyl sulfate (SDS).
6. Micro-spin columns (GE Healthcare).

2.2.7 Protein Purification

1. Gel filtration column such as Sephadex G-25, BioGel P-30, or similar with the appropriate molecular weight cutoff (GE LifeSciences).
2. 100 mM phosphate-buffered saline between pH 7.0 and 7.5.

2.2.8 Single-Molecule Immobilization

1. Spectroscopic grade acetone and methanol (Sigma Ltd) and acetic acid.
2. Pegylation buffer: 10 mL ultrapure water + 84 mg sodium bicarbonate.

3. Aminopropyl silane (Sigma), Biotinylated-PEG, and mPEG (SusTech, Darmstadt, Germany) (*see* **Note 5**).
4. Streptavidin was purchased from Invitrogen and dissolved at a concentration 5 mg/mL in 50 mM Tris–HCl (pH 8.1) and 50 mM NaCl. The solution should be stored at 4 °C immediately after preparation.
5. Imaging buffer: 50 mM Tris–HCl (pH 8.1), 6 % (w/w) glucose, 1 % 2-mercaptoethanol, 0.1 mg/mL glucose oxidase type II-S from Aspergillus niger (Sigma), 0.02 mg/mL glucose catalase (Roche Diagnostics) (*see* **Notes 6** and **7**).
6. Quartz microscope slides 1 in. × 3 in. × 1 mm (Finkenbeiner, Waltham) for prism-type Sm-TIR.
7. Fluorescence beads for instrument alignment and to generate the mapping algorithm that allows correlating individual donor spots with individual acceptor spots can be purchased from Molecular Probes (FluoSpheres carboxylated, 0.2 μm). Commercial stock solution should be diluted 1000-fold in 5 mM HCl before injecting on the quartz slide (*see* **Note 8**).

3 Methods

Independent of whether the FRET assay is going to be performed under steady-state, time-resolved, or single-molecule regime, the first steps involve the labeling and purification of the biomolecules under investigation. We assume at this stage that the reader has identified an appropriate FRET pair for the particular question to be addressed. We will start with a description of general methodologies for protein dye-labeling, post-synthetic fluorescent labeling and purification of DNA, and we will continue with a general description of the protocol to follow under each experimental regime.

3.1 Preparation of Fluorescently Labeled DNA

1. Dissolve the amine-modified oligonucleotide in 100 μL of doubly distilled water.
2. Precipitate overnight at −20 °C by adding 0.1 volume of NaOAc (3 M) and 2.5 volumes (250 μL) of 100 % ethanol (*see* **Note 9**).
3. After centrifugation of the solution at ~16,000 × *g* for 30 min, carefully discard the supernatant and perform a quick spin to remove remaining ethanol. Air-dry pellets on bench or under vacuum.
4. Resuspend pellet in water to achieve a final concentration of 25 μg/μL. The solution is stable at −20 °C for long periods. Oligonucleotide concentration can be obtained from the absorption spectrum assuming that an optical density of 1.0

corresponds to a DNA concentration of 33 μg/mL in a 1 cm path cell.

5. Dissolve the commercial amine-reactive dye (commonly a succinimidyl ester derivative of the dye) in DMSO to reach a concentration of 250 μg/14 μL. This concentration of dye is optimized to label 100 μg of DNA. Remaining stock of dye in DMSO should be dried under vacuum until next use.
6. Mix 4 μL of the DNA strand to be labeled with the amine-reactive dye (250 μg/14 μL) in a total volume of 100 μL of labeling buffer.
7. Incubate the reaction overnight at room temperature with gentle shaking, particularly during the first hour.
8. Precipitate the reaction mixture overnight at –20 °C as described in **step 2**. Centrifuge at ~16,000×*g* for 30 min and carefully remove the supernatant. Redissolve the pellet in 100 μL of 50 % formamide and purify as described in Subheadings 3.2 and 3.3.

3.2 Purification of Labeled DNA by Polyacrylamide Gel Electrophoresis

1. To purify the fluorescently labeled DNA, pour a 1.5 mm thick polyacrylamide gel. Acrylamide percentage should be chosen according to the DNA length. In this laboratory we normally use 20 % for oligonucleotides less than 25 bases, 15 % between 25 and 40 bases, and 10 % for longer oligonucleotides. For a 10 % gel we mix 30 mL of 20 % (w/v) acrylamide (7 M urea) with another 30 mL of 7 M urea/1× TBE. Adding 450 μL of ammonium persulfate (10 % w/v) and 45 μL of TEMED will start the polymerization reaction. Leave the gel for at least an hour to ensure complete polymerization.
2. Because urea accumulates in the wells, before pre-running the gel for 1 h at 18 W, wash the wells with 1× TBE running buffer using a syringe.
3. Incubate the formamide solution containing the labeled DNA for 2 min at 90 °C to disrupt any secondary structure.
4. Just before loading samples, re-wash loading wells with a syringe. Use xylene cyanol FF and bromophenol blue in 100 % formamide as size markers. On a 10 % gel, xylene cyanol and bromophenol dyes migrate approximately as oligonucleotides of 55 and 10 nucleotides, respectively (*see* **Note 10**).
5. Run the gel until the dyes have migrated a minimum of 2/3 of the plate length. Remove glass plates and carefully put the acrylamide gel on a DarkReader transilluminator (DR45M, VWR International) to visualize DNA bands by fluorescence.

Cut the bands containing the proper length of fluorescent DNA and also the bands corresponding to xylene cyanol or bromophenol blue as control for electroelution. Samples can either be stored in a freezer or used immediately for electroelution.

3.3 DNA Recovery

3.3.1 Recovery by Electroelution

We described here the protocol to be followed when using a Harvard-type electroeluter. Different steps of this protocol might need to be adapted to each specific electroeluter model.

1. Fill the electroeluter apparatus with 0.25× TBE and run at 200 V for 30 min to clean it. Discard the running buffer, fill the electroluter tank with freshly made 0.25× TBE buffer, and add 200 μL of 8 M ammonium acetate in the trapping wells.
2. Carefully place the acrylamide bands in small pieces in the recipient wells and as an indication of the electroelution progress place also a marker band in a separate well. Run the electroluter at 120 V for 1 h or until the maker band is completely free of dye.
3. Block the trapping wells using tips cut at ~5 mm length and remove the buffer from the electroluter tank using a syringe.
4. Collect the fluorescent DNA from each trapping well in separate Eppendorfs and wash each well with running buffer to collect residual DNA sample left.
5. Fill each Eppendorf with 100 % ethanol and store overnight at −20 °C for precipitation.
6. Resuspend the pellet in appropriate buffer (i.e., 50 mM Tris–HCl, pH 8.0) and calculate the labeling efficiency using an absorption spectrophotometer.

3.3.2 Recovery by Crush and Soak

1. For crush and soak, place the acrylamide band cut in small pieces in a micro-spin column (GE Healthcare) and fill with buffer to cover all the acrylamide pieces. Seal the column and leave it overnight with gentle shaking.
2. Cut the sealed end of the column, place the column inside an Eppendorf, and centrifuge at ~16,000 × *g* for 2 min. Add 1/10 volume of 3 M NaOAc and proceed as in Subheading 3.3.1, **steps 5** and **6**.

3.4 Preparation of Fluorescently Labeled Protein

1. Dissolve the protein at 50–100 μM in 10 mM Tris–HCl or phosphate buffer at pH 7.0–7.5 at room temperature.
2. Prepare a 1 mM stock solution of the maleimide dye in DMF or DMSO. Protect all dye stock solutions from light as much as possible.
3. Add the maleimide reagent dropwise to the protein solution to give approximately 10 mol of reagent per mole of protein.
4. Allow the reaction to proceed for 2 h at room temperature or overnight at 4 °C (*see* **Note 11**).

3.5 Purification of Fluorescently Labeled Protein

1. Sephadex G-25 or similar gel filtration media of the appropriate molecular weight cutoff can be used to separate the protein-dye conjugate from the unreacted labeling reagent. For proteins

larger than 5000 Mr we normally use PD-10 columns (GE Healthcare) with excellent separation.

2. Cut the sealed bottom of the PD-10 column and fill up the column with 25 mL of equilibration buffer (i.e., 50 mM Tris–HCl, pH 7.5) and discard the flow-through.
3. Apply the protein sample to the column (maximum 2.5 mL). If the protein sample to be purified is less than 2.5 mL, add equilibration buffer to adjust the total volume to 2.5 mL. Let the sample pass through and discard the flow through.
4. Elute the protein with 3.5 mL of equilibration buffer and collect each fraction in a test tube or an Eppendorf. The first eluted fluorescent band will correspond to the dye-protein conjugate. Labeled samples should be stored under the same conditions as the unlabeled protein, but protecting the stock as much as possible from light.
5. Calculate the degree of labeling using the expression

$$\frac{\text{moles dye}}{\text{moles protein}} = \frac{A_{\text{dye}}}{\varepsilon_{\text{dye}}} \times \frac{\text{MW protein}}{\text{mg protein / mL}}$$

3.6 Steady-State FRET Experiments

Steady-state FRET experiments can reveal information about protein-DNA interactions by monitoring (1) the increase in FRET signal as a consequence of formation of the protein-DNA complex using an intermolecular FRET assay (Fig. 1a), (2) the change in intramolecular FRET signal (Fig. 1b) as a result of a binding-induced conformational change either in the protein or in the DNA, and (3) the breakdown of the inter- or intramolecular FRET upon protein-induced DNA cleavage (Fig. 3). The following protocols describe the basis to perform these assays in a quantitative manner.

3.6.1 Protein-DNA Binding Assay

1. Load 120 μL of a nanomolar substrate solution in a clean quartz microcell and place it in the sample chamber of the fluorimeter, thermostated at the appropriate temperature. In an intermolecular-type assay, the substrate DNA solution will contain only the donor fluorophore, whilst in an intramolecular-type assay it will contain a donor–acceptor labeled construct (*see* **Note 12**).
2. Titrate the substrate solution with increasing concentrations of binding partner by adding aliquots of stock solution. For each addition, allow to equilibrate for 5 min and take two emission fluorescence spectra, one exciting at the maximum of the donor and a second one at a wavelength that excites only the acceptor species. We usually collect the fluorescence spectra with 8 nm slit-width and 10 nm/s acquisition rate (*see* **Note 13**).

3. If the substrate is labeled with donor and acceptor, an equivalent construct carrying only the donor species should be prepared and used to independently collect the emission fluorescence spectrum of the donor in the absence of acceptor.
4. Normalize the donor-only spectrum obtained in **step 3** to the maximum of the donor fluorescence band in each FRET spectrum in the titration and subtract it from the latter (Fig. 2d).
5. The *ratioA* parameter is proportional to the amount of acceptor fluorescence emission due to energy transfer (Fig. 2d) and can be calculated for each data point in the titration by first normalizing the donor-only fluorescence spectrum to the maximum of the donor band in the FRET experimental spectrum, and then by subtracting the normalized donor spectrum from the experimental FRET spectra. The resulting fluorescence curve contains only acceptor signal. The *ratioA* is then obtained following the expression

$$ratioA = \frac{F_{acc}^{excD}}{F_{acc}^{excA}}$$

where F_{acc}^{excD} represents the fluorescence spectrum of the acceptor obtained from the experimental FRET spectrum after subtracting the normalized donor contribution, and F_{acc}^{excA} represents the fluorescence spectrum of the acceptor obtained by exciting at its absorption maximum or at a wavelength where only the acceptor absorbs.

6. The *ratioA* parameter can then be converted in energy transfer efficiency values by applying

$$E_{FRET} = \frac{ratioA - \frac{A_{acc}^{excD}}{A_{acc}^{excA}}}{\frac{A_{don}^{excD}}{A_{acc}^{excA}}}$$

where A_{acc}^{excD} represents the absorption of the acceptor at the donor excitation wavelength (i.e., for fluorescein, 490 nm), A_{acc}^{excA} represents the absorption of the acceptor at its excitation wavelength (i.e., for Cy3, 547 nm), and A_{don}^{excD} represents the absorption of the donor at its excitation wavelength (*see* **Note 14**).

3.6.2 Protein-DNA Cleavage Assay

1. Proceed as in **step 1** of the previous Subheading 3.6.1, but instead of collecting the emission of the donor and acceptor as a function of the wavelength, measure a time-trace by exciting the donor and collecting simultaneously the emission of the

donor and acceptor against time, by continuously switching the monochromator between both wavelengths.

2. Before the enzyme is added to the cell, the relative intensities of both signals and therefore the FRET efficiency should be constant. Then add the enzyme and mix manually with the pipette. Interaction between the substrate and the enzyme will be reported as an increase in the donor signal with a concomitant exponential decrease in FRET efficiency. Monitor the cleavage reaction until both signals reach a plateau.
3. Fit the measured exponential increase of the donor signal or the exponential decrease of the FRET efficiency to an equation of the type $F(t) = F_o(t) + Ae^{-}kt$, where A represents the amplitude and k is the pseudo-first-order rate constant.

3.7 Time-Resolved FRET Experiments

In tr-FRET, the fluorescence lifetime of the donor is measured in the absence and in the presence of the acceptor fluorophore as a function of increasing concentrations of interacting partners (Fig. 3b, c). In binding assays, the difference in the lifetimes in the presence and absence of acceptor is used to obtain, for each data point in the titration, the Gaussian distribution of distances that represent the populations present (Eq. 8). In cleavage assays, either the average lifetime (Fig. 3d) or the amplitude of substrate and product (Fig. 3d inset) is plotted against time to extract the pseudo-first order cleavage rate constant. Here, we describe the protocols for these types of experiments using time-resolved FRET based on the TCSPC technique.

3.7.1 Time-Resolved Protein-DNA Binding Assays

1. Place a scattering solution in the time-resolved fluorimeter (i.e., colloidal silica, diluted milk) and measure the instrument response function with approximately 10,000 counts peak intensity and with the excitation and emission wavelengths at the excitation wavelength of the donor. Adjust the detection frequency, so that the maximum number of counts does not exceed the ~5 % of the laser repetition rate to avoid counting artifacts.
2. Place a quartz cell of donor-only DNA construct and measure the fluorescence decay up to 10,000 counts peak intensity.
3. Add increasing amounts of acceptor construct (intermolecular assay) or binding partner (intramolecular assay). After each addition, allow to equilibrate for ~5 min and collect the fluorescence decay at the wavelength of the maximum of the donor fluorescence spectrum.
4. Use the instrument response function previously obtained to deconvolute all fluorescence decays.
5. Fit the deconvoluted donor-only decay to Eq. 7 or 8 depending on the number of expected lifetime distributions.

6. Fit the deconvoluted donor-acceptor decay functions to the Gaussian distribution of distances represented by Eq. 8 and plot their relative percentual contributions as a function of added species (*see* **Note 15**).

3.7.2 Time-Resolved Protein-Induced DNA Cleavage Assays

1. Proceed as described in **steps 1** and **2** of Subheading 3.7.1.
2. Set up the time-resolved fluorimeter to collect consecutive fluorescence decays at the maximum emission wavelength of the donor for the estimated duration of the cleavage reaction. Each donor decay should be collected in time mode in which the fluorescence is collected for a constant period of time. For example, collect continuous decays at ½ min intervals for a 1 h reaction. Automatically save each of them in a separate file.
3. Add the enzyme and quickly start the acquisition.
4. Use the instrument response function previously obtained to deconvolute all fluorescence decays.
5. Fit the deconvoluted donor-only decay to Eq. 6 or to a sum of exponential decays.
6. Fit the deconvoluted donor-acceptor decay to a sum of exponentials (i.e., donor-only lifetime that represents the cleavage product plus donor lifetime in the presence of acceptor that represents the substrate).
7. To extract kinetic information it is possible to plot the average lifetime defined as shown in Eq. 9 or the time evolution of the pre-exponential factors associated to the product and the substrate.

3.8 Single-Molecule FRET

In this section we describe only those methods related to the measurement of single-molecule FRET between proteins and DNA using the total internal reflection technique.

3.8.1 Quartz Slide Cleaning Protocol

Independent of whether the quartz slide and the glass cover slip are new or recycled, cleaning of the slides and cover slip is an absolutely crucial step for single-molecule applications. In this section we provide the protocol that the authors commonly use in their laboratory.

1. Sonicate quartz slide and cover slip in (1) 20 % detergent solution for 15 min, (2) water for 5 min, (3) acetone for 15 min, (4) water for 5 min, (5) 1 M KOH for 15 min, (6) methanol for 15 min, (7) 1 M KOH for 15 min, and (8) water 5 min.
2. Dry with nitrogen or compressed air.
3. Quartz slide and cover slip are passed through a torch flame to remove impurities and moisture.

3.8.2 Quartz Slide Pegylation Protocol

1. Pour 100 mL of methanol in a reaction flask and add 5 mL of concentrated acetic acid with a glass pipette.
2. Add 1 mL of aminopropyl silane using a glass pipette and mix well.
3. Pour the mixture in the slide/cover slip container and incubate for 10 min.
4. Sonicate the reaction container for 2 min and then incubate for another 10 min at room temperature.
5. Rinse slides and/or cover slip with methanol and ultrapure water.
6. Dry with nitrogen or compressed air and put them in clean tip boxes with water at the bottom.
7. For the PEG coating of five slides, take 1–2 mg of biotinylated PEG and 80 mg of mPEG and put them in 1.5 mL Eppendorf tube.
8. Add 320 μL of pegylation buffer and mix gently with a pipette. Spin it for 1 min at 13,000 × *g*.
9. Drop 70 μL of the mPEG/biotin-PEG mixture on each slide.
10. Place very gently a cover slip on the top avoiding the formation of bubbles.
11. Allow them to incubate for 2 h at room temperature in a dark and flat place. After ~10 min of placing the cover slip on the top of the slide, check, and restore any misplaced cover slip.
12. Remove the cover slip from the quartz slide, rinse them with ultrapure water, and dry using a flow of nitrogen or air (*see* **Note 16**).
13. Form a channel using double sticky tape sandwiched between the slide and the cover slip with a ~5 mm gap between the two tape strips (*see* **Note 17**).
14. Add water to the channel and look in the microscope to see if the channel is clean before adding the immobilization reagents.
15. Dilute 25-fold the initial streptavidin stock solution in 50 mM Tris–HCl (pH 8.1) and 25 mM NaCl and add 50 μL to the slide channel.
16. Allow another 10 min for streptavidin binding.
17. Wash unbound streptavidin with 60 μL of 50 mM Tris–HCl (pH 8.1) and 25 mM NaCl.
18. Add 60 μL of a 50–100 pM biotinylated biomolecule in 50 mM Tris–HCl (pH 8.1) and 25 mM NaCl and allow 5 min for binding to the streptavidin-coated surface.
19. Look for the density and quality of labeling in the Sm-TIR microscope and if adequate proceed to add the imaging buffer. If the amount of immobilized molecules is below the desired level repeat **steps 18** and **19** (*see* **Note 18**).

3.8.3 Data Collection and Analysis

1. Prior to collecting single-molecule data, a mapping algorithm that correlates donor spots (left half of the EMCCD camera) with their acceptor counterparts (right half) needs to be created. In the author's laboratory, this is performed using a solution of 200 nm fluorescence beads and a program written in IDL v. 6 software (ITT Visual Information Systems, USA).
2. Fluorescence data at donor and acceptor wavelengths are acquired from single molecules by using total internal reflection fluorescence microscopy with 532 nm laser excitation.
3. Data are acquired using a laboratory-written Visual C++ v. 6 program with integration times ranging from 16 to 100 ms depending on the sample dynamics. Measurements are performed at room temperature.
4. Single-molecule FRET efficiency after background correction is approximated by $(I_A/[I_A + I_D])$, where I_A and I_D are the fluorescence intensities of the acceptor and donor, respectively. Because the quantum yields and detection efficiencies of Cy3 and Cy5 are very close, E_{app} (the apparent FRET value observed) closely matches the true efficiency of energy transfer. Data analysis is performed using laboratory-written analysis routines developed in MATLAB 7 (The MathWorks Inc., USA).
5. Single-molecule FRET histograms are obtained by averaging the first ten frames of each FRET trace for every individual molecule after manually filtering photobleaching and blinking effects. States are identified from E_{app} histograms, and dwell times are analyzed only if the time resolution allows the clear observation of transitions (more than five data points per dwell time). Rapidly fluctuating molecules undergo more transitions than slowly fluctuating ones, and thus in order to avoid bias toward fast rates, dwell time histograms are obtained by using a weighting factor inversely proportional to the number of transitions observed for each molecule. These dwell time histograms are then fitted to a single-exponential function to obtain the lifetimes of each state, the inverse of the rate of conversion. For the heterogeneity analysis, the average of the dwell times is calculated for each state for each individual molecule (*see* **Note 19**).

4 Notes

1. Depending on the type of Sm-TIR setup, an oil-immersion or a water-immersion objective would be required. For prism-type sm-TIR, as the one described in Fig. 4a, b, because the evanescence wave is generated in the interface quartz-slide/water, a water immersion objective is the appropriate choice due to its higher penetration depth. However, for the objective-type

TIR, the evanescence wave generates at the bottom interface (cover slip/water) and thus a higher NA oil-immersion objective that has better performance can be used. In this case it is important to use fused silica cover slides to reduce background from glass luminescence.

2. It has been shown that linear PEG interacts with unfolded proteins preventing refolding [46]. Groll et al. [53] overcame this problem using branched star-shaped PEG polymers (SusTech, Darmstadt, Germany). Following this alternative, they were able to observe reversible folding and unfolding steps from single-protein molecules.
3. As a general rule, all solutions involved in sample preparation should be tested in the single-molecule equipment. A common source of impurities is water, and thus quartz-bidistilled water is recommended but it should be monitored regularly for contaminants. In the author's laboratory, ultrapure fresh commercial water (Sigma or other general chemical supplier) is always available for testing and sample preparation.
4. This solution is kept at 4 °C in the dark. Keep this solution only for 2 or 3 weeks for better efficiency.
5. Once aminopropyl silane has been used, the remaining stock can be reused but care must be taken to store it properly. It should be dehydrated in vacuum for ~15 min, sealed under nitrogen atmosphere, and kept at −20 °C.
6. Trolox (6-hydroxy-2,7,5,8)-tetramethylchroman-2-carboxylic acid) is a water-soluble derivative of vitamin E that has recently been proposed as a good alternative to the triplet-state quencher 2-mercaptoethanol, significantly reducing undesirable Cy5 blinking events.
7. Care should be taken when using higher glucose concentrations as the increase in solution viscosity could artificially affect the intrinsic dynamics of the biomolecule or complex under investigation.
8. To achieve a homogeneous distribution of fluorescence beads it is convenient to sonicate the solution to be injected on the slides for 5 min. This will substantially decrease bead aggregation. Quartz slides coated with fluorescence beads can be active for long periods of time by sealing them with epoxy resin.
9. In those cases where the DNA is not easily precipitated from solution, add glycogen to a final 0.05–1 μg/μL concentration. Use of up to 1 μL of glycogen per 20 μL of the solution, overnight incubation, and storage at −20 °C help to increase the recovery yields.
10. Both xylene cyanol FF and bromophenol blue dyes show UV absorption when visualized on a TLC plate (Merck, Silica gel plate F254 20×20 cm). It is thus preferable not to load the

dyes in the same wells as the DNA oligonucleotides since it could complicate visualization and contaminate the DNA.

11. After the bioconjugation reaction has reached the desired time, an excess of glutathione or 2-mercaptoethanol may be added to consume the remaining thiol-reactive reagent so that no reactive species are present during the purification step.
12. When using intermolecular FRET to measure protein-DNA binding affinities with a donor-labeled DNA substrate and an acceptor-labeled protein, it is advisable to choose a FRET pair with an absorption spectral separation as large as possible (i.e., Cy3/Cy5). Thus, direct excitation of the acceptor-labeled protein during the titration is minimized. This is particularly crucial in those systems with poor affinities.
13. Before recording a FRET titration with many data points, it is advisable to check that the fluorescence intensity at any stage of the titration does not saturate the photomultiplier. This can be done using a quick test at the starting and end points of the titration to find the optimum instrumental parameters (slit width and photomultiplier voltage).
14. When calculating the FRET efficiency following the enhanced acceptor emission as described, it is assumed that the acceptor quantum yield is independent of the excitation wavelength, which is normally the case. However, if the shape of an acceptor emission spectra, obtained after subtraction of the normalized donor spectrum from the FRET spectrum, does not correspond to a standard acceptor spectrum, it is clear that both donor and acceptor fluorophores are strongly coupled. In these conditions, energy transfer can indeed take place but it is not clear if Eqs. 1 and 2 represent a good estimation of the physical process, suggesting that other mechanisms of energy transfer should be considered.
15. When analyzing the data from tr-FRET, it is usually convenient in order to get highly accurate data to use global fitting procedures instead of fitting each fluorescence decay individually. Thus, lifetimes and amplitudes are the best possible values to reproduce all experimental data.
16. Pegylated quartz slides and cover slips can be stored for some time in a dark and dry place (~1 week depending on the laboratory conditions). In our laboratory, we use black-tape-wrapped Corning tubes to avoid light and plastic film to avoid moisture.
17. This defines a volume that is approximately 5 mm wide, 18 mm long, and 100 μm tall. Due to the low volume of solvent required to fill the channel, evaporation is usually quick reducing long-term use of the sample. To avoid this, two procedures can be followed: (1) sealing completely the channel with fast-drying time epoxy resin and (2) using a diamond drill (Eternal

Tools, UK), make two ~0.75 mm holes on opposite sides of the quartz slide, so that they can be used to inject sample and additives in real-time at any stage of the experiment. The sides of the quartz slide not covered by the double-sticky tape are sealed with epoxy resin.

18. To be sure that the fluorescence observed in the CCD camera is arising from biomolecules specifically attached to the quartz slide perform the following test: (1) Expose a certain region of the slide to high intensity laser light to promote complete photobleaching and switch off the laser light for 2 min. (2) Switch back on the laser light and if molecules can again be observed; it is very likely that the sample contains either free molecules in solution or adsorption/desorption processes from nonspecifically attached molecules are taking place.
19. Single-molecule FRET histograms usually show a "zero peak" Gaussian distribution caused mainly by incomplete labeling (lack of acceptor) or fast photobleaching during the first few frames of the experiment. Thus, care should be taken to (1) work at very low laser intensities and (2) to ensure that the lowest value of "true" FRET that can be expected from the biomolecule behavior is well separated from the "zero peak." Usually, with a 10 % acceptor leakage in the donor channel, low FRET values higher than ~0.15 can be safely analyzed as arising from an active FRET state of the biomolecule.

Acknowledgments

We thank the Biological and Biotechnology Science Research Council (UK), the Royal Society (UK), and the National Sciences and Engineering Research Council (Canada) and the Universities of Sherbrooke (Canada) and St Andrews (UK) for financial support. We also thank all members of our labs for helpful discussion and critical reading of the manuscript. J.C.P. is a Fellow of the Scottish Universities Physics Alliance (SUPA).

References

1. Hillisch A, Lorenz M, Diekmann S (2001) Recent advances in FRET: distance determination in protein-DNA complexes. Curr Opin Struct Biol 11:201–207
2. Holbrook SR (2005) RNA structure: the long and the short of it. Curr Opin Struct Biol 15:302–308
3. Yan Y, Marriott G (2003) Analysis of protein interactions using fluorescence technologies. Curr Opin Chem Biol 7:635–640
4. Michalet X, Kapanidis AN, Laurence T, Pinaud F, Doose S, Pflughoefft M, Weiss S (2003) The power and prospects of fluorescence microscopies and spectroscopies. Annu Rev Biophys Biomol Struct 32:161–182
5. Lorenz M, Hillisch A, Payet D, Buttinelli M, Travers A, Diekmann S (1999) DNA bending induced by high mobility group proteins studied by fluorescence resonance energy transfer. Biochemistry 38:12150–12158

6. Selvin PR (2000) The renaissance of fluorescence resonance energy transfer. Nat Struct Biol 7:730–734
7. Stuhmeier F, Hillisch A, Clegg RM, Diekman S (2000) Fluorescence energy transfer analysis of DNA structures containing several bulges and their interaction with CAP. J Mol Biol 302:1081–1100
8. Clegg RM (1992) Fluorescence resonance energy transfer and nucleic acids. Methods Enzymol 211:353–388
9. Stryer L, Haugland RP (1967) Energy transfer: a spectroscopic ruler. Proc Natl Acad Sci U S A 58:719–726
10. Stuhmeier F, Hillisch A, Clegg RM, Diekman S (2000) Practical aspects of fluorescence resonance energy transfer (FRET) and its applications in nucleic acid biochemistry. In: Travers A, Buckle M (eds) DNA-protein interactions. Oxford University Press, Oxford, pp 77–94
11. Bera A, Roche AC, Nandi PK (2007) Bending and unwinding of nucleic acid by prion protein. Biochemistry 46:1320–1328
12. Lorenz M, Diekmann S (2006) Distance determination in protein-DNA complexes using fluorescence resonance energy transfer. Methods Mol Biol 335:243–255
13. Passner JM, Steitz TA (1997) The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer. Proc Natl Acad Sci U S A 94:2843–2847
14. Hieb AR, Halsey WA, Betterton MD, Perkins TT, Kugel JF, Goodrich JA (2007) TFIIA changes the conformation of the DNA in TBP/TATA complexes and increases their kinetic stability. J Mol Biol 372:619–632
15. Dragan AI, Klass J, Read C, Churchill ME, Crane-Robinson C, Privalov PL (2003) DNA binding of a non-sequence-specific HMG-D protein is entropy driven with a substantial non-electrostatic contribution. J Mol Biol 331:795–813
16. Kuznetsov SV, Sugimura S, Vivas P, Crothers DM, Ansari A (2006) Direct observation of DNA bending/unbending kinetics in complex with DNA-bending protein IHF. Proc Natl Acad Sci U S A 103:18515–18520
17. Lorenz M, Hillisch A, Goodman SD, Diekmann S (1999) Global structure similarities of intact and nicked DNA complexed with IHF measured in solution by fluorescence resonance energy transfer. Nucleic Acids Res 27:4619–4625
18. Chapados BR, Hosfield DJ, Han S, Qiu J, Yelent B, Shen B, Tainer JA (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. Cell 116:39–50
19. Xiao J, Singleton SF (2002) Elucidating a key intermediate in homologous DNA strand exchange: structural characterization of the RecA-triple-stranded DNA complex using fluorescence resonance energy transfer. J Mol Biol 320:529–558
20. McKinney SA, Joo C, Ha T (2006) Analysis of single-molecule FRET trajectories using hidden Markov modeling. Biophys J 91:1941–1951
21. Gupta RC, Golub EI, Wold MS, Radding CM (1998) Polarity of DNA strand exchange promoted by recombination proteins of the RecA family. Proc Natl Acad Sci U S A 95:9843–9848
22. Kuznetsov SV, Kozlov AG, Lohman TM, Ansari A (2006) Microsecond dynamics of protein-DNA interactions: direct observation of the wrapping/unwrapping kinetics of single-stranded DNA around the E. coli SSB tetramer. J Mol Biol 359:55–65
23. Lucius AL, Jason Wong C, Lohman TM (2004) Fluorescence stopped-flow studies of single turnover kinetics of E. coli RecBCD helicase-catalyzed DNA unwinding. J Mol Biol 339:731–750
24. Kvaratskhelia M, Wardleworth BN, Bond CS, Fogg JM, Lilley DM, White MF (2002) Holliday junction resolution is modulated by archaeal chromatin components in vitro. J Biol Chem 277:2992–2996
25. Furey WS, Joyce CM, Osborne MA, Klenerman D, Peliska JA, Balasubramanian S (1998) Use of fluorescence resonance energy transfer to investigate the conformation of DNA substrates bound to the Klenow fragment. Biochemistry 37:2979–2990
26. Mukhopadhyay J, Mekler V, Kortkhonjia E, Kapanidis AN, Ebright YW, Ebright RH (2003) Fluorescence resonance energy transfer (FRET) in analysis of transcription-complex structure and function. Methods Enzymol 371:144–159
27. Heyduk T, Niedziela-Majka A (2001) Fluorescence resonance energy transfer analysis of Escherichia coli RNA polymerase and polymerase-DNA complexes. Biopolymers 61: 201–213
28. Margeat E, Kapanidis AN, Tinnefeld P, Wang Y, Mukhopadhyay J, Ebright RH, Weiss S (2006) Direct observation of abortive initiation and promoter escape within single immobilized transcription complexes. Biophys J 90: 1419–1431
29. Kapanidis AN, Margeat E, Ho SO, Kortkhonjia E, Weiss S, Ebright RH (2006) Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. Science 314: 1144–1147

30. Lee SP, Han MK (1997) Fluorescence assays for DNA cleavage. Methods Enzymol 278:343–363
31. Eggeling C, Jager S, Winkler D, Kask P (2005) Comparison of different fluorescence fluctuation methods for their use in FRET assays: monitoring a protease reaction. Curr Pharm Biotechnol 6:351–371
32. Ghosh SS, Eis PS, Blumeyer K, Fearon K, Millar DP (1994) Real time kinetics of restriction endonuclease cleavage monitored by fluorescence resonance energy transfer. Nucleic Acids Res 22:3155–3159
33. Hiller DA, Rodriguez AM, Perona JJ (2005) Non-cognate enzyme-DNA complex: structural and kinetic analysis of EcoRV endonuclease bound to the EcoRI recognition site GAATTC. J Mol Biol 354:121–136
34. Ray PC, Fortner A, Darbha GK (2006) Gold nanoparticle based FRET assay for the detection of DNA cleavage. J Phys Chem B 110:20745–20748
35. Lin J, Gao H, Schallhorn KA, Harris RM, Cao W, Ke PC (2007) Lesion recognition and cleavage by endonuclease V: a single-molecule study. Biochemistry 46:7132–7137
36. van der Meer BW (2002) Kappa-squared: from nuisance to new sense. J Biotechnol 82: 181–196
37. Klostermeier D, Millar DP (2001) Time-resolved fluorescence resonance energy transfer: a versatile tool for the analysis of nucleic acids. Biopolymers 61:159–179
38. Cornish PV, Ha T (2007) A survey of single-molecule techniques in chemical biology. ACS Chem Biol 2:53–61
39. Ha T (2001) Single-molecule fluorescence resonance energy transfer. Methods 25:78–86
40. Ha T (2004) Structural dynamics and processing of nucleic acids revealed by single-molecule spectroscopy. Biochemistry 43:4055–4063
41. Ritort F (2006) Single-molecule experiments in biological physics: methods and applications. J Phys Condens Matter 18:R531–R583
42. Moerner WE, Fromm DP (2003) Methods of single-molecule fluorescence spectroscopy and microscopy. Rev Sci Instrum 74:3597
43. Haustein E, Schwille P (2007) Fluorescence correlation spectroscopy: novel variations of an established technique. Annu Rev Biophys Biomol Struct 36:151–169
44. Schwille P (2003) TIR-FCS: staying on the surface can sometimes be better. Biophys J 85:2783–2784
45. Wazawa T, Ueda M (2005) Total internal reflection fluorescence microscopy in single molecule nanobioscience. Adv Biochem Eng Biotechnol 95:77–106
46. Rasnik I, McKinney SA, Ha T (2005) Surfaces and orientations: much to FRET about? Acc Chem Res 38:542–548
47. Cisse I, Okumus B, Joo C, Ha T (2007) Fueling protein DNA interactions inside porous nanocontainers. Proc Natl Acad Sci U S A 104:12646–12650
48. Myong S, Bruno MM, Pyle AM, Ha T (2007) Spring-loaded mechanism of DNA unwinding by hepatitis C virus NS3 helicase. Science 317:513–516
49. Lu HP, Iakoucheva LM, Ackerman EJ (2001) Single-molecule conformational dynamics of fluctuating noncovalent DNA-protein interactions in DNA damage recognition. J Am Chem Soc 123:9184–9185
50. Segers-Nolten GM, Wyman C, Wijgers N, Vermeulen W, Lenferink AT, Hoeijmakers JH, Greve J, Otto C (2002) Scanning confocal fluorescence microscopy for single molecule analysis of nucleotide excision repair complexes. Nucleic Acids Res 30:4720–4727
51. Lemay JF, Penedo JC, Tremblay R, Lilley DM, Lafontaine DA (2006) Folding of the adenine riboswitch. Chem Biol 13:857–868
52. Braslavsky I, Hebert B, Kartalov E, Quake SR (2003) Sequence information can be obtained from single DNA molecules. Proc Natl Acad Sci U S A 100:3960–3964
53. Groll J, Amirgoulova EV, Ameringer T, Heyes CD, Rocker C, Nienhaus GU, Moller M (2004) Biofunctionalized, ultrathin coatings of cross-linked star-shaped poly(ethylene oxide) allow reversible folding of immobilized proteins. J Am Chem Soc 126:4234–4239
54. Adachi K, Yasuda R, Noji H, Itoh H, Harada Y, Yoshida M, Kinosita K Jr (2000) Stepping rotation of F1-ATPase visualized through angle-resolved single-fluorophore imaging. Proc Natl Acad Sci U S A 97:7243–7247
55. Kastner CN, Prummer M, Sick B, Renn A, Wild UP, Dimroth P (2003) The citrate carrier CitS probed by single-molecule fluorescence spectroscopy. Biophys J 84:1651–1659
56. Boukobza E, Sonnenfeld A, Haran G (2001) Immobilization in surface-tethered lipid vesicles as a new tool for single biomolecule spectroscopy. J Phys Chem B 105:12165–12170
57. Kapanidis AN, Weiss S (2002) Fluorescent probes and bioconjugation chemistries for single-molecule fluorescence analysis of biomolecules. J Chem Phys 117:10953–10964
58. Sapsford KE, Berti L, Medintz IL (2006) Materials for fluorescence resonance energy transfer analysis: beyond traditional donor-acceptor combinations. Angew Chem Int Ed Engl 45:4562–4589

Chapter 9

Precise Identification of Genome-Wide Transcription Start Sites in Bacteria by 5′-Rapid Amplification of cDNA Ends (5′-RACE)

Dominick Matteau and Sébastien Rodrigue

Abstract

Transcription start sites are commonly used to locate promoter elements in bacterial genomes. TSS were previously studied one gene at a time, often through 5′-rapid amplification of cDNA ends (5′-RACE). This technique has now been adapted for high-throughput sequencing and can be used to precisely identify TSS in a genome-wide fashion for practically any bacterium, which greatly contributes to our understanding of gene regulatory networks in microorganisms.

Key words Transcription, Promoter, RNA polymerase holoenzyme, Transcription start site, Rapid amplification of cDNA ends, 5′-RACE, Genome wide, Next-generation sequencing, RNA-seq, Transcriptomics

1 Introduction

Bacteria adapt to a changing environment by rapidly modulating the expression of several genes. The first step in this process occurs through transcription, a process in which RNA polymerase produces an RNA molecule from a DNA template. The core of the bacterial RNA polymerase is composed of five subunits, $\beta\beta'\alpha_2\omega$, and adopts a crab-claw structure sufficient for transcription but not for recognition of promoters and transcription initiation [1, 2]. In bacteria, initiation of transcription is orchestrated by a holoenzyme complex composed of the RNA polymerase bound by one σ factor [3]. *Escherichia coli* expresses seven different σ factors competing for the RNA polymerase core to form distinct functional holoenzymes that can initiate transcription at specific promoters. To initiate transcription, the holoenzyme complex interacts with two short DNA sequences of approximately 6 base pairs (bp) respectively located approximately 10 and 35 bp upstream from the transcription start site (TSS) [1, 4–6]. If present, two other

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_9,

DNA elements can also affect holoenzyme binding affinity to promoters: the UP element, a motif located ~20 bp upstream of the −35 element and recognized by the C-terminal domains of the RNA polymerase α subunits, and the extended −10 element, a 3–4 bp motif located immediately upstream of the −10 element that is recognized by the RNA polymerase σ subunit [1, 4–6]. Once bound to a promoter, the holoenzyme forms the open complex, in which a short segment of promoter DNA around the TSS is unwound and the template strand is inserted into the active site of the polymerase [1, 2, 7]. Then, the synthesis of the DNA template-directed RNA chain begins at the TSS, a position marked by a triphosphate nucleoside at the 5′ end of the nascent RNA transcript.

Identifying TSS is important because their upstream region contains promoter elements that directly contribute to transcription regulation, thereby providing insights on how regulatory networks are organized in the cell. Additionally, TSS allows the determination of the 5′ untranslated region (UTR) of transcripts; it can be used to define operons and can lead to the identification of small noncoding RNAs. Initially, bacterial TSS have been identified using two experimental procedures: primer extension [8] and S1 nuclease protection mapping assays [9]. Generally laborious, requiring large amount of bacterial RNA, and displaying variable sensitivity, these methods have been progressively replaced by PCR-based techniques commonly called rapid amplification of cDNA ends (RACE) [10–12]. Throughout the years, the initial RACE strategy has been optimized considerably by several laboratories, thus improving many aspects of the procedure such as reducing amplification of prematurely terminated first-strand cDNAs, improving amplification specificity, and adding the capacity of cloning and sequencing 3′ ends of cDNAs [13–16]. However, these approaches require separate experiments to identify TSS of each individual gene, and characterizing all the TSS of an entire genome would be extremely challenging considering the throughput of these techniques. More recently, with the advent of functional genomics, microarray technology offered new opportunities to characterize bacterial gene expression and TSS at an unprecedented genome-wide level and rapidly became the standard for genome-wide transcriptome analyses of prokaryotes [17, 18]. Even if significant improvements in microarray technology have been made throughout the years, thereby leading to outstanding breakthroughs in transcriptomics studies [19–22], important limitations such as variable background noise and sensitivity, smaller dynamic range, as well as lower resolution and coverage of large genomes led to the development of a totally new approach to study transcriptomes [17, 18]. Nowadays, methods based on next-generation sequencing mostly replaced microarray-based approaches to study transcriptomes, principally because

RNA sequencing (RNA-seq) provides the same information given by microarrays but with greater resolution and sensitivity, and also bypass the limitation of detecting only predefined regions [23]. Since its first utilization, RNA-seq technology improved significantly, and modifications in library preparation procedure offer the possibility to accurately map TSS and transcription termination sites (genome-wide 5′- and 3′-RACE), operons, 5′- and 3′-UTRs, as well as antisense RNAs and small noncoding RNAs [24–27]. Transcriptomes can thus be studied with an unprecedented level of details.

Although genome-wide 5′-RACE is commonly used in combination with RNA-seq, the following protocol focuses on the former method and is summarized in Fig. 1. Briefly, bacterial cells are lysed and total RNA is purified. The purified RNA quality is then verified by Bioanalyzer to ensure good integrity of mRNAs and thus proper characterization of TSS (*see* Fig. 2). RNA is then fragmented by magnesium-catalyzed hydrolysis, and the 5′-RACE library is prepared. mRNA molecules that do not contain a triphosphorylated 5′ extremity are depleted using the XRN-1 5′ → 3′ exoribonuclease. At the end of the procedure, cDNA is synthesized from processed RNAs ligated to 5′ and 3′ adapters, and library is amplified by qPCR. Library quality and concentration are assessed by Bioanalyzer (*see* Fig. 2) before high-throughput sequencing. 5′-RACE reads are aligned on the desired reference genome and can be visualized using IGV [28] or the UCSC Genome Browser [29] to identify TSS at specific locations (*see* Fig. 3).

2 Materials

2.1 Bacterial Culture and RNA Purification

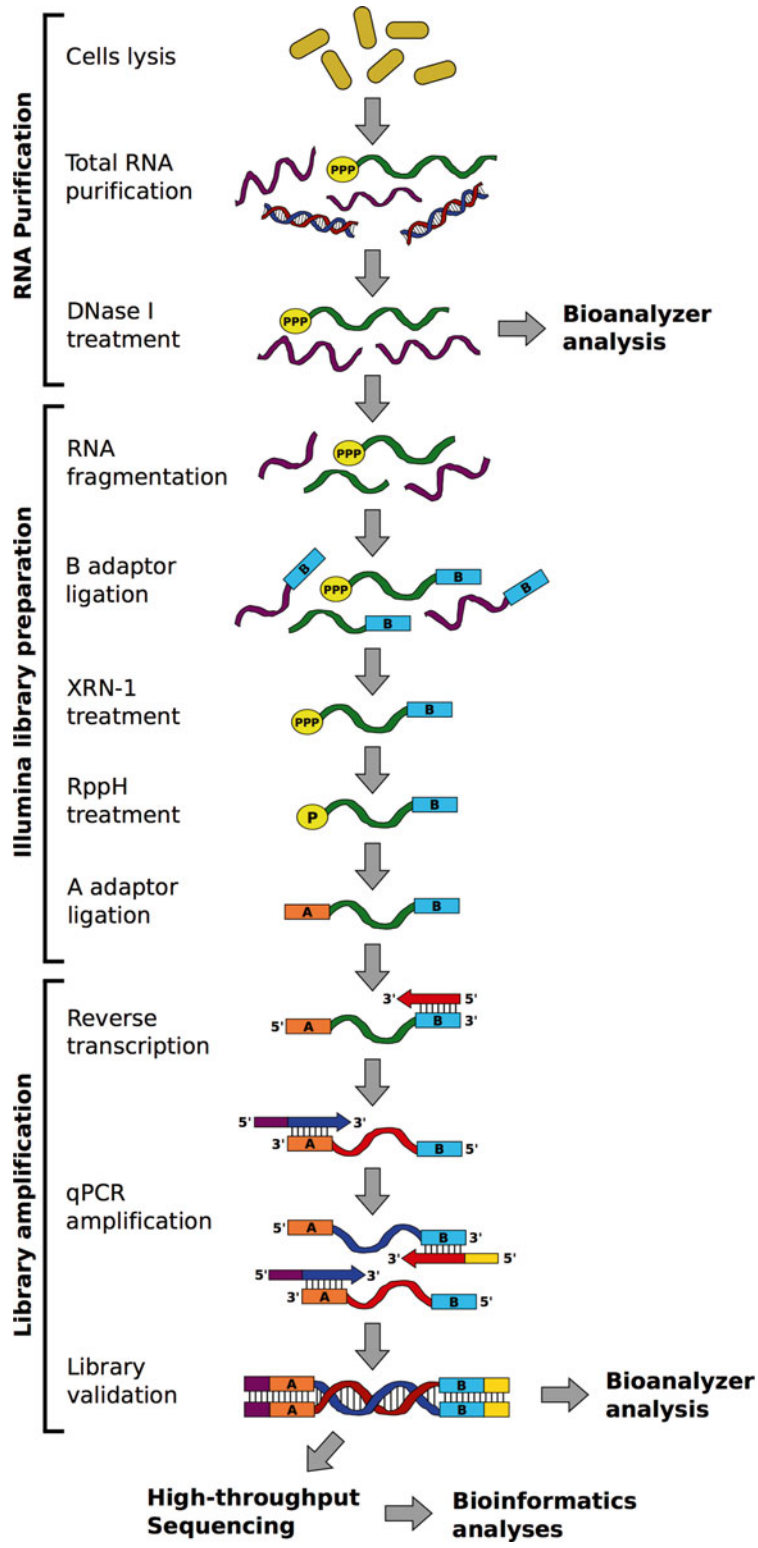
1. LB broth, sterilized by autoclave.
2. Shaking incubator.
3. 1.5 ml microcentrifuge tubes.
4. QIAzol Lysis Reagent (QIAGEN, 79306) or equivalent.
5. Vortex mixer.
6. Direct-zol RNA MiniPrep Kit (Zymo Research, R2052).
7. Molecular grade sterile H_2O.
8. Nuclease-free 0.2 ml PCR tubes.
9. RNase-free DNase I and supplied 10× reaction buffer.
10. RNA Clean and Concentrator-5 Kit (Zymo Research, R1016).
11. 2100 Bioanalyzer instrument and RNA 6000 Nano Kit (Agilent Technologies, 5067-1511).

2.2 Illumina Library Preparation

1. Thermocycler.
2. Nuclease-free 0.2 ml PCR tubes.

3. Molecular grade sterile H_2O.
4. RNA fragmentation buffer 5×: 200 mM Tris-acetate pH 8.1, 500 mM potassium acetate, 150 mM magnesium acetate. Sterilize by autoclave.
5. RNA Clean and Concentrator-5 Kit (Zymo Research, R1016).
6. 3′-B adaptor (*see* Table 1).
7. Molecular grade 100 % DMSO.
8. T4 RNA Ligase 1 and supplied 10× reaction buffer (Enzymatics, 20,000 U/ml, L6050L).
9. RNase Inhibitor (Enzymatics, 40,000 U/ml, Y9240L).
10. XRN-1 (New England Biolabs, 1,000 U/ml, M0338L).
11. 10× NEBuffer 3 (New England Biolabs, B7003S).
12. RppH and supplied 10× NEBuffer 2 (New England Biolabs, 5,000 U/ml, M0356S).
13. 5′-A adaptor (*see* Table 1).
14. Agencourt RNAClean XP SPRI beads (Beckman Coulter, A63987).
15. TruSeq-DSN-R oligo (*see* Table 1).
16. Molecular grade 10 mM dNTPs.
17. M-MuLV Reverse Transcriptase and supplied 10× reaction buffer (Enzymatics, 200,000 U/ml, P7040L).
18. Actinomycin D 2.5 μg/μl.

Fig. 1 Overview of the 5′-RACE methodology. (1) For RNA purification, bacterial cells are lysed using a phenol/guanidine lysis reagent, and total RNA is purified by column purification. Contaminating DNA molecules are then eliminated by a DNase I treatment and by a second RNA purification. Quality of purified RNA is then verified by Bioanalyzer to ensure good integrity of mRNAs and thus proper characterization of transcription start sites. (2) For Illumina library preparation, RNA is fragmented by magnesium-catalyzed hydrolysis, and an Illumina DNA-RNA hybrid adaptor (represented by the letter B, 3′-B adaptor; *see* Table 1) is ligated using T4 RNA Ligase to the 3′ end of fragmented RNA molecules. Then, mRNA molecules that contain transcription start sites (TSS) are enriched from other RNAs using XRN-1, an 5′ → 3′ exoribonuclease that does not process RNA with a triphosphorylated 5′ end. After, mRNAs are treated with an RNA 5′ pyrophosphohydrolase (RppH) to remove a pyrophosphate from the 5′ end of triphosphorylated RNAs to leave mRNAs with a monophosphate. A second Illumina DNA-RNA hybrid adaptor (represented by the letter A, 5′-A adaptor; *see* Table 1) is ligated to the monophosphorylated 5′ end of mRNAs, immediately next to potential TSS. (3) For library amplification, cDNA is synthesized from RNA containing 3′ and 5′ adaptors using a reverse transcriptase and an oligonucleotide (TruSeq-DSN-R; *see* Table 1) annealing on the B adaptor. 5′-RACE library is finally amplified by qPCR with a forward oligonucleotide (IGA-PCR-PE-F; *see* Table 1) annealing on the A adaptor and a reverse oligonucleotide (TruSeq-MPEX-R; *see* Table 1) annealing on the B adaptor. Both oligonucleotides contain a 5′ anchor sequence used for Illumina sequencing, and the reverse oligonucleotide contains an index of six nucleotides specific for each sample multiplexed for the sequencing procedure


RNA Purification
Cells lysis
Total RNA purification
DNase I treatment
Bioanalyzer analysis
Illumina library preparation
RNA fragmentation
B adaptor ligation
XRN-1 treatment
RppH treatment
A adaptor ligation
Library amplification
Reverse transcription
qPCR amplification
Library validation
Bioanalyzer analysis
High-throughput Sequencing
Bioinformatics analyses

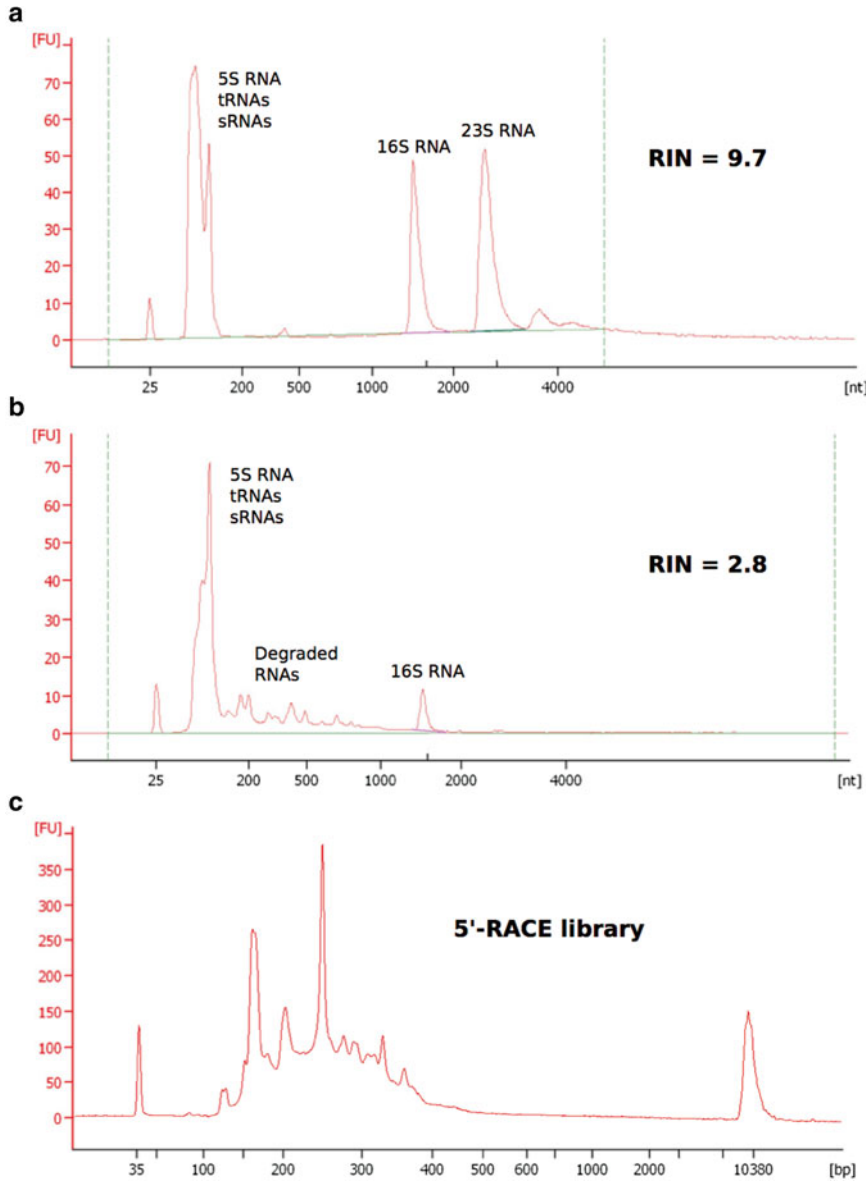


Fig. 2 Bioanalyzer profiles of total RNA extractions and 5′-RACE library. (**a**) Typical Bioanalyzer profile of a bacterial total RNA extraction of good quality. 16S and 23S ribosomal RNAs are visible through narrow peaks at approximately 1,500 and 2,900 nucleotides (nt), respectively. Other peaks, generally present from 50 to 200 nt, can also be visible and represent 5S ribosomal RNA, tRNAs, as well as small RNAs (sRNAs). These small peaks are especially noted when using phenol/guanidine extraction methods. The Bioanalyzer software automatically calculates an RNA integrity number (RIN) representing the quality of RNA extraction. An RIN close to 10 indicates intact RNA extraction, while an RIN close to 1 indicates a degraded and thus RNA extraction of poor quality. (**b**) Bioanalyzer profile of a heavily degraded RNA extraction, characterized by the decrease or absence of 16S and 23S ribosomal RNA signals, a modification in their abundance ratio, as well as the presence of a smear between 50 and 1,000 nt. This type of profile is associated with a low RIN value, which indicates that the RNA extraction is not suitable for a genome-wide characterization of transcription start sites (TSS) by 5′-RACE methodology. (**c**) Bioanalyzer profile of a typical 5′-RACE library. Due to the heterogeneous size distribution of mRNA molecules as well as the constant position of targeted TSS, the 5′-RACE Illumina library does not follow a typical normal distribution usually obtained in RNA sequencing or DNA sequencing

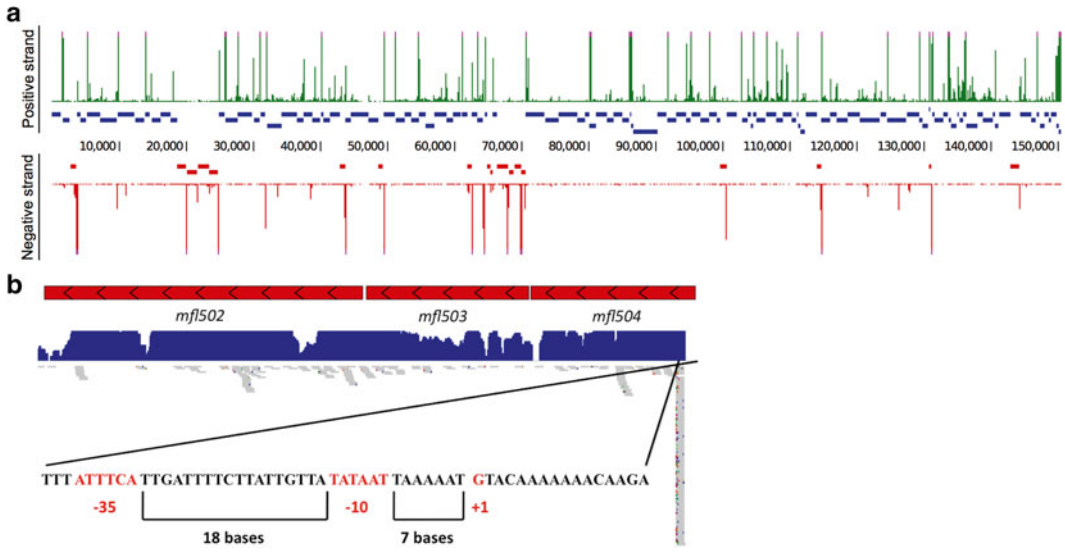


Fig. 3 Example of results generated by 5′-RACE sequencing. (**a**) Genome-wide characterization of transcription start sites (TSS) of the near-minimal bacterium *Mesoplasma florum* following 5′-RACE sequencing. The first and the last tracks show 5′-RACE signals aligned on the first 150 kb of *M. florum*'s genome on the positive DNA strand (*green*) and the negative DNA strand (*red*), respectively. *Pink dots* at the top of peaks indicate signals beyond the represented *y*-axis maximal value. Second and third tracks show predicted genes on the positive DNA strand (*blue*) and the negative DNA strand (*red*), respectively. Position across the represented portion of the genome is indicated in base pairs. (**b**) Promoter characterization of three genes (first track, *red*) apparently transcribed from an operon starting at *mfl504*. The second track shows transcription level of the represented genes obtained by RNA sequencing (*blue*). The third track shows 5′-RACE reads aligned at the represented genomic locus (*gray*). A clear TSS can be determined at the beginning of *mfl504*, along with a typical −10 promoter element recognized by principal σ factors in bacteria

Table 1
Oligonucleotide sequences used for 5′-RACE Illumina library preparation

Name	Nucleotide sequence (5′–3′)
3′-B adaptor[a,b]	/5Phos/rArGrArUrCrGrGrArArGAGCACACGTCT/3AmMO/
5′-A adaptor[b]	ACACGACGrCrUrCrUrUrCrCrGrArUrCrU
TruSeq-DSN-R	AGACGTGTGCTCTTCCGATCT
IGA-PCR-PE-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCT CTTCCGATCT
TruSeq-MPEX-R[c]	CAAGCAGAAGACGGCATACGAGAT-INDEX-GTGACTGGAGTTCAGACG TGTGCTCTTCCGATC

[a]Oligonucleotide modification: /5Phos/, 5′ phosphorylation; /3AmMO/, 3′ amino modifier
[b]r corresponds to RNA bases
[c]-INDEX- corresponds to a specific combination of 6 bp for each sample multiplexed in the library

19. Agencourt AMPure XP SPRI beads (Beckman Coulter, A63881).
20. IGA-PCR-PE-F and TruSeq-MPEX-R oligos (*see* Table 1).
21. VeraSeq 2.0 High-Fidelity DNA polymerase and supplied 5× reaction buffer (Enzymatics, 2,000 U/ml, P7511L).
22. SYBR Green I 10×.
23. Real-time quantitative PCR system.
24. 2100 Bioanalyzer instrument and High Sensitivity DNA Analysis Kit (Agilent Technologies, 5067-4626).

3 Methods

The following protocol was designed to work with any bacterium but was tested on exponential growth phase *E. coli* as well as *Mesoplasma florum* cultures. The 5′-RACE sequencing library preparation procedure was tested from total RNA extractions, but could be conducted on rRNA-depleted RNA samples to minimize rRNA read contamination and increase genome-wide TSS coverage. rRNA can be depleted by different methods and commercially available kits such as Evrogen's duplex-specific nuclease (DSN), Ribo-Zero from Epicentre, or MICROBExpress from Ambion [30–35]. It is also important to note that the described library preparation and sequencing are designed and tested using Illumina technology, but should provide equivalent results with any next-generation sequencing platform. An overview of the 5′-RACE protocol is illustrated in Fig. 1.

3.1 Bacterial Culture

1. Inoculate 5 ml of desired growth medium containing appropriate antibiotics with 50 µl of a pre-culture of the desired bacterial strain (1/100 dilution).
2. Grow cells in a shaking incubator at the appropriate temperature until the culture reaches the desired optical density (*see* **Note 1**).
3. Transfer ~1.5 ml of the culture into a 1.5 ml microcentrifuge tube and centrifuge at max speed to pellet cells.
4. Remove supernatant and repeat the last step (*see* **Note 2**).
5. Remove supernatant and proceed immediately to RNA purification.

3.2 Total RNA Purification

1. Add 800 µl of QIAzol Lysis Reagent (or equivalent) and vortex at high speed for about 30 s.
2. Incubate at room temperature for about 10 min and vortex well every 1–2 min to completely lyse cells. No cell pellet should be visible before proceeding to the next step.

3. Centrifuge at 13,000 × *g* for 2 min to pellet non-soluble material and recover supernatant.
4. Purify total RNA with the Direct-zol RNA MiniPrep Kit (Zymo Research) according to the manufacturer's specifications.
5. Elute RNA with 25 µl of molecular grade H_2O in a 1.5 ml microcentrifuge tube.

3.3 DNase I Treatment and Bioanalyzer Analysis

1. Mix the following reagents in a 0.2 ml PCR tube to remove possible DNA contamination:

   21 µl purified total RNA from the previous step.

   2.5 µl 10× DNase I reaction buffer.

   1.5 µl DNase I.

   Total: 25 µl.
2. Incubate at 37 °C for about 20 min.
3. Purify DNase I-treated RNA with the RNA Clean and Concentrator-5 Kit (Zymo Research) according to the manufacturer's specifications.
4. Elute RNA with 25 µl of molecular grade H_2O in a 1.5 ml microcentrifuge tube.
5. Keep ~2 µl of purified total RNA on ice for sample analysis and store the remaining volume at −80 °C to prevent RNA degradation.
6. Evaluate RNA quality and concentration on a 2100 Bioanalyzer instrument using an RNA 6000 Nano Kit (Agilent Technologies; *see* **Note 3** and Fig. 2).

3.4 RNA Fragmentation for 5′-RACE Illumina Library Preparation

1. Mix the following reagents in a 0.2 ml PCR tube to perform magnesium-catalyzed hydrolysis:

   2 µg previously frozen total RNA from **step 5** of Subheading 3.3 (maximum 20 µl).

   5 µl RNA fragmentation buffer 5×.

   Complete to 25 µl with molecular grade H_2O.
2. Incubate at 75 °C for 10 min in a thermocycler and immediately put on ice.
3. Purify using the RNA Clean and Concentrator-5 Kit (Zymo Research) and elute in 10 µl of molecular grade H_2O.

3.5 3′-B Adaptor Ligation

1. Mix the following reagents in a 0.2 ml PCR tube to prepare an RNA ligation reaction:

   10 µl fragmented RNA from the previous step.

   4 µl 10 µM 3′-B adaptor (*see* Table 1 and **Note 4**).

   2 µl 10× T4 RNA Ligase buffer.

2 μl 100 % molecular grade DMSO.

Total: 18 μl.

2. Incubate at 65 °C for 1 min in a thermocycler and put immediately on ice.
3. Add the following reagents directly to the tube from the previous step:

   0.5 μl RNase Inhibitor.

   1.5 μl T4 RNA Ligase.

   Total: 20 μl.
4. Incubate at 37 °C for 60 min in a thermocycler.
5. Purify using the RNA Clean and Concentrator-5 Kit (Zymo Research) and elute in 14 μl of molecular grade H_2O.

3.6 Degradation of 5′ Monophosphorylated RNAs by XRN-1

1. Mix the following reagents in a 0.2 ml PCR tube to prepare the XRN-1 reaction:

   14 μl 3′-B adaptor-ligated RNA from the last step.

   2 μl 10× NEBuffer 3.

   0.5 μl RNase Inhibitor.

   0.5 μl molecular grade H_2O.

   3 μl XRN-1 exoribonuclease.

   Total: 20 μl.
2. Incubate at 37 °C for 60 min in a thermocycler.
3. Purify using the RNA Clean and Concentrator-5 Kit (Zymo Research) and elute in 16 μl of molecular grade H_2O.

3.7 Pyrophosphate Removal of 5′ Triphosphorylated RNAs by RppH

1. Mix the following reagents in a 0.2 ml PCR tube to prepare the RppH reaction:

   16 μl XRN-1-treated RNA from the last step.

   2 μl 10× NEBuffer 2.

   1.0 μl RNase Inhibitor.

   1.0 μl RppH.

   Total: 20 μl.
2. Incubate at 37 °C for 30 min in a thermocycler.
3. Purify using the RNA Clean and Concentrator-5 Kit (Zymo Research) and elute in 10 μl of molecular grade H_2O.

3.8 5′-A Adaptor Ligation

1. Mix the following reagents in a 0.2 ml PCR tube to prepare an RNA ligation reaction:

   10 μl fragmented RNA from the last step.

   4 μl 10 μM 5′-A adaptor (*see* Table 1 and **Note 4**).

   2 μl 10× T4 RNA Ligase buffer.

2 μl 100 % molecular grade DMSO.

Total: 18 μl.

2. Incubate at 65 °C for 1 min in a thermocycler and put immediately on ice.
3. Add the following reagents directly to the tube from the previous step:

   0.5 μl RNase Inhibitor.

   1.5 μl T4 RNA Ligase.

   Total: 20 μl.
4. Incubate at 37 °C for 60 min in a thermocycler.
5. Purify RNA by adding and mixing 36 μl of RNA SPRI magnetic beads (RNAClean XP) to the solution (1.8:1 volume ratio; *see* **Note 5**).
6. Incubate for 5 min at room temperature.
7. Place the reaction tube on a 0.2 ml magnetic support and wait for 2–3 min until all beads are immobilized on the tube wall.
8. Remove supernatant carefully while keeping the tube on the magnetic support.
9. Add 150 μl of fresh 70 % ethanol, wait approximately 30 s, and remove the liquid while keeping the tube on the magnetic support.
10. Repeat **step 7** once and remove any trace of remaining 70 % ethanol using a 1–10 μl tip.
11. Incubate beads at room temperature for 10–15 min to dry residual ethanol (*see* **Note 6**).
12. Remove the tube from the magnetic support and resuspend beads in 11 μl of molecular grade H_2O by pipetting thoroughly but carefully.
13. Incubate at room temperature for at least 1 min.
14. Place the tube on the magnetic support again and wait for 1–2 min until all beads are immobilized on the tube wall.
15. Transfer 10 μl of the supernatant to a new 0.2 ml PCR tube (leave 1 μl of the supernatant in the tube to avoid bead contamination).

3.9 Reverse Transcription

1. Mix the following reagents in a 0.2 ml PCR tube to prepare a reverse transcription reaction:

   10 μl 5′-A- and 3′-B-ligated RNA from the last step.

   1 μl 10 μM TruSeq-DSN-R oligo (*see* Table 1).

   2 μl 10 mM dNTPs.

   Total: 13 μl.

2. Incubate at 65 °C for 2 min in a thermocycler and rapidly quench on ice.
3. Add the following reagents directly to the tube from the previous step:

   2 µl 10× M-MuLV RT reaction buffer.

   1.4 µl molecular grade H_2O.

   1 µl RNase Inhibitor.

   1 µl M-MuLV Reverse Transcriptase.

   1.6 µl 2.5 µg/µl actinomycin D.

   Total: 20 µl.
4. Mix well, spin down, and incubate in a thermocycler using the following program:

   42 °C—10 min.

   42–50 °C—40 min (increment of 0.2 °C/min).

   50 °C—10 min.

   85 °C—10 min (heat inactivation).
5. Purify cDNA by adding and mixing 36 µl of DNA SPRI magnetic beads (AMPure XP) to the solution (1.8:1 volume ratio; *see* **Note 5**).
6. Repeat **steps 6–14** of Subheading 3.8, except that 80 % ethanol is used instead of 70 % and that DNA elution is done in 16 µl of molecular grade H_2O.
7. Transfer 15 µl of the supernatant to a new 0.2 ml PCR tube (leave 1 µl of the supernatant in the tube to avoid bead contamination).

3.10 5′-RACE Library Amplification

1. Prepare two reaction tubes (amplification duplicates) with the following reagents to amplify the 5′-RACE library by qPCR:

   2 µl of purified cDNA from the last step.

   14.6 µl molecular grade H_2O.

   5 µl 5× VeraSeq reaction buffer.

   1 µl IGA-PCR-PE-F oligo (*see* Table 1).

   1 µl 10 µM TruSeq-MPEX-R oligo (contains index; *see* Table 1).

   0.5 µl 10 mM dNTPs.

   0.63 µl 10× SYBR Green I.

   0.25 µl VeraSeq 2.0 DNA polymerase.

   Total: 25 µl (×2 reactions).

2. Mix well, spin down, and amplify DNA by qPCR using the following program:

Denature at 98 °C for 30 s	
98 °C—15 s	
60 °C—15 s	
72 °C—15 s	Repeat amplification cycle (*n*) times.
Final extension at 72 °C for 2 min.	

3. Stop amplification reaction cycles during the late exponential phase and proceed to the final extension step (*see* **Note 7**).
4. Pool the two qPCR reactions (50 μl) and purify DNA by adding and mixing 45 μl of DNA SPRI magnetic beads (AMPure XP) to the solution (0.9:1 volume ratio; *see* **Note 5**).
5. Repeat **steps 6–14** of Subheading 3.8, except that DNA elution is done in 16 μl of molecular grade H_2O.
6. Transfer 15 μl of the supernatant to a new 0.2 ml PCR tube (leave 1 μl of the supernatant in the tube to avoid bead contamination).
7. Evaluate library size and concentration using a 2100 Bioanalyzer instrument with a High Sensitivity DNA Analysis Kit (Agilent Technologies; *see* **Note 8** and Fig. 2).

3.11 DNA Sequencing and Data Analysis

1. If more than one sample is prepared, mix indexed libraries to obtain sufficient coverage of the corresponding reference genome(s) (*see* **Note 9**).
2. Perform single-end or paired-end sequencing on multiplexed samples at your Illumina sequencing service provider.
3. If needed, separate Illumina reads (demultiplexing) with NovoBarcode (www.novocraft.com) based on the index used for library amplification (TruSeq-MPEX-R oligo; *see* Table 1).
4. Verify quality of raw sequence data obtained from the sequencing step using the FastQC quality control program (www.bioinformatics.babraham.ac.uk/projects/fastqc/).
5. Align forward read of mate pair on reference genome(s) using the Burrows-Wheeler Aligner (BWA) [36] or equivalent.
6. Verify mapping quality using SAMStat [37].
7. Filter alignment using SAMtools view [38] to discard aligned reads with a quality score below 10.
8. Sort alignment using SAMtools sort [38].

9. To calculate 5′-RACE 1 bp-read density, treat aligned reads to conserve only the first nucleotide of each read and then calculate density with bedtools genomecov [39].
10. Visualize BedGraph files generated by bedtools with IGV [28] or with UCSC Genome Browser [29] (*see* Fig. 3).

4 Notes

1. Cells can be harvested at different optical densities depending on experimental needs. Generally, harvesting bacterial cells during the mid-exponential growth phase yields good results.
2. Volume of culture used for RNA extraction has to be modified depending on the organism type and cell concentration of the culture. Refer to the Direct-zol RNA MiniPrep Kit (Zymo Research, R2052) for further details. In our experience, 3 ml of exponential growth phase *E. coli* (~3×10^8 cells at OD600 of 0.5–0.6) yields to good-quality total RNA preparations.
3. For qualitative and quantitative analysis using the 2100 Bioanalyzer instrument with the RNA 6000 Nano Kit, RNA samples must have a concentration between 25 and 500 ng/μl. To avoid exceeding maximal concentration, samples can be first quantified by NanoDrop and diluted accordingly.
4. DNA-RNA hybrid oligos should be stored at −80 °C to avoid degradation.
5. Always use RNA or DNA SPRI magnetic beads tempered at room temperature for optimal binding efficiency. Modifying the SPRI volume ratio affects bead capacity to bind small DNA or RNA fragments [40]. For DNA, a SPRI volume ratio of 1.8:1 allows the recovery of fragments larger than 80 bp, while a ratio of 1:1 allows the recovery of fragments larger than 100 bp.
6. It is important to let beads dry completely to avoid ethanol contamination and maximize RNA or DNA elution efficiency. Generally, waiting for 10–15 min at room temperature is sufficient to eliminate residual ethanol, but beads should not be overdried since this could also lead to poor recovery.
7. The number of PCR amplification cycles depends on the amount of starting material. Generally, between 12 and 20 cycles are enough to reach the late exponential amplification phase. It is very important to stop the qPCR amplification reaction during the 72 °C step (extension) to keep DNA double stranded.
8. After amplification, libraries often have a concentration between 10 and 30 ng/μl and must be diluted 1/20 for

Bioanalyzer analysis using a High Sensitivity DNA assay chip. Alternatively, size distribution can be evaluated by agarose gel electrophoresis, and DNA quantification can be obtained using a nanodrop.

9. The coverage value used for a 5′-RACE experiment is generally around 50×, but can be modified according to the desired sequencing depth and experimental needs. Sample coverage is calculated as follows:

   (Sequencing read length in bp × theoretical number of reads)/Genome size in bp.

   The theoretical number of reads for one sample is calculated by multiplying the total number of reads normally obtained with the chosen sequencing technology by the sample pooling ratio (ng of sample in mix/total ng of pooled mix).

Acknowledgment

We thank Pierre-Étienne Jacques for technical assistance, Thomas Knight Jr. for *M. florum* genome annotations, the *Centre de calcul scientifique* of Université de Sherbrooke for computational resources and technical support, and Vincent Baby for its precious comments on the manuscript. This work was supported by the *Fonds québécois de la recherche sur la nature et les technologies* through a MSc scholarship awarded to D.M. and a Projet de recherche en équipe grant awarded to S.R. and Vincent Burrus. S.R. holds a Chercheur-Boursier Junior 1 award from the Fonds de recherche Québec-Santé.

References

1. Browning DF, Busby SJ (2004) The regulation of bacterial transcription initiation. Nat Rev Microbiol 2:57–65
2. Lee DJ, Minchin SD, Busby SJW (2012) Activating transcription in bacteria. Annu Rev Microbiol 66:125–152
3. Gruber TM, Gross CA (2003) Multiple sigma subunits and the partitioning of bacterial transcription space. Annu Rev Microbiol 57:441–466
4. Murakami KS, Masuda S, Campbell EA et al (2002) Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. Science 296:1285–1290
5. Campbell EA, Muzzin O, Chlenov M et al (2002) Structure of the bacterial RNA polymerase promoter specificity σ subunit. Mol Cell 9:527–539
6. Lonetto M, Gribskov M, Gross CA (1992) The Sigma70 family : sequence conservation and evolutionary relationships. J Bacteriol 174: 3843–3849
7. Zupancic M, Record TM (1998) RNA polymerase-promoter interactions : the comings and goings of RNA polymerase. J Bacteriol 180:3019–3025
8. Thompson J, Radonovich M, Salzman N (1979) Characterization of the 5′-terminal structure of simian virus 40 early mRNA's. J Virol 31:437–446
9. Berk A, Sharp P (1977) Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids. Cell 12:721–732
10. Frohman MA, Dush MK, Martin GR (1988) Rapid production of full-length cDNAs from

rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proc Natl Acad Sci 85:8998–9002

11. Ohara O, Dorit R, Gilbert W (1989) One-sided polymerase chain reaction: the amplification of cDNA. Proc Natl Acad Sci U S A 86:5673–5677
12. Loh E, Elliott J, Cwirla S et al (1989) Polymerase chain reaction with single-sided specificity: analysis of t cell receptor delta chain. Science 2:8–11
13. Frohman M (1994) On beyond classic RACE (rapid amplification of cDNA ends). PCR Methods Appl 4:S40–S58
14. Schaefer BC (1995) Revolutions in rapid amplification of cDNA ends: new strategies for polymerase chain reaction cloning of full-length cDNA ends. Anal Biochem 227:255–273
15. Li Z, Yu M, Zhang H et al (2005) Improved rapid amplification of cDNA ends (RACE) for mapping both the 5′ and 3′ terminal sequences of paramyxovirus genomes. J Virol Methods 130:154–156
16. Tillett D, Burns BP, Neilan BA (2000) Optimized rapid amplification of cDNA ends (RACE) for mapping bacterial mRNA transcripts. Biotechniques 28:448–456
17. Siezen RJ, Wilson G, Todt T (2010) Prokaryotic whole-transcriptome analysis: deep sequencing and tiling arrays. J Microbial Biotechnol 3:125–130
18. van Vliet AHM (2010) Next generation sequencing of microbial transcriptomes: challenges and opportunities. FEMS Microbiol Lett 302:1–7
19. Selinger DW, Cheung KJ, Mei R et al (2000) RNA expression analysis using a 30 base pair resolution Escherichia coli genome array. Nat Biotechnol 18:1262–1268
20. McGrath PT, Lee H, Zhang L et al (2007) High-throughput identification of transcription start sites, conserved promoter motifs and predicted regulons. Nat Biotechnol 25: 584–592
21. Rasmussen S, Nielsen HB, Jarmer H (2009) The transcriptionally active regions in the genome of Bacillus subtilis. Mol Microbiol 73:1043–1057
22. Toledo-Arana A, Dussurget O, Nikitas G et al (2009) The Listeria transcriptional landscape from saprophytism to virulence. Nature 459:950–956
23. Hoen PAC, Ariyurek Y, Thygesen HH et al (2008) Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms. Nucleic Acids Res 36, e141
24. Thomason MK, Bischler T, Eisenbart SK et al (2015) Global transcriptional start site mapping using differential RNA sequencing reveals novel antisense RNAs in Escherichia coli. J Bacteriol 197:18–28
25. Wade JT, Grainger DC (2014) Pervasive transcription: illuminating the dark matter of bacterial transcriptomes. Nat Rev Microbiol 12:647–653
26. Shinhara A, Matsui M, Hiraoka K et al (2011) Deep sequencing reveals as-yet-undiscovered small RNAs in *Escherichia coli*. BMC Genomics 12:428
27. Conway T, Creecy J, Maddox S (2014) Unprecedented high-resolution view of bacterial operon architecture revealed by RNA sequencing. MBio 5:1–12
28. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192
29. Kent WJ, Sugnet CW, Furey TS et al (2002) The human genome browser at UCSC. Genome Res 12:996–1006
30. Bogdanova E, Shagina I, Mudrik E et al (2009) DSN depletion is a simple method to remove selected transcripts from cDNA populations. Mol Biotechnol 41:247–253
31. Bogdanova E, Shagina I, Yanushevich YG et al (2011) Preparation of prokaryotic cDNA for full-scale transcriptome analysis. Russ J Bioorg Chem 37:775–778
32. Vandernoot V, Langevin S, Solberg OD et al (2012) cDNA normalization by hydroxyapatite chromatography to enrich transcriptome diversity in RNA-seq applications. Biotechniques 53:373–380
33. Yi H, Cho Y-J, Won S et al (2011) Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. Nucleic Acids Res 39, e140
34. He S, Wurtzel O, Singh K et al (2010) Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. Nat Methods 7:807–812
35. Giannoukos G, Ciulla DM, Huang K et al (2012) Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. Genome Biol 13:R23
36. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760

37. Lassmann T, Hayashizaki Y, Daub CO (2011) SAMStat: monitoring biases in next generation sequencing data. Bioinformatics 27:130–131
38. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
39. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842
40. Rodrigue S, Materna AC, Timberlake SC et al (2010) Unlocking short read sequencing for metagenomics. PLoS One 5, e11840

Chapter 10

Analysis of DNA Supercoiling Induced by DNA–Protein Interactions

David J. Clark and Benoît P. Leblanc

Abstract

Certain DNA-interacting proteins induce a pronounced bending in the double helix and cause topological stresses that are compensated by the formation of supercoils in DNA. Such supercoils, when forming on a circular plasmid, give rise to a series of topoisomers that run at different speeds during electrophoresis. The number of supercoils introduced in the plasmid can provide information on the protein; it can, for example, help determine the number of nucleosomes that are assembled on the plasmid or indicate whether the DNA-bending activity of a transcription factor is important enough to cause a topological stress. Because a DNA–protein activity can lead to either an overwinding or an underwinding of the helix, supercoiling can occur in either direction. Determining whether a plasmid contains positively or negatively supercoiled DNA is possible, thanks to an agarose gel containing an intercalating agent known to positively supercoil DNA, such as chloroquine. The speed of migration of the topoisomers varies in a characteristic way in the presence and absence of the agent. Topoisomer standards can furthermore be generated to allow the easy evaluation of the number of supercoils induced in a plasmid by a DNA–protein interaction.

Key words DNA–protein interaction, Supercoiling, Topology, Chloroquine

1 Introduction

The two strands of the DNA helix are unable to move around each other freely without the help of a nicking activity. Their overwinding or underwinding therefore results in the accumulation of a topological stress that must be somehow dissipated or compensated, as is easily visualized in an overwound rubber band that seeks to snap back into its original conformation or starts forming supercoils.

Certain DNA-binding proteins lead to a pronounced bending in the axis of the double helix. This is the case with histone octamers at the core of nucleosomes [1], as well as with nonhistone proteins of the HMGB family [2] and with certain transcription factors containing an HMG-box motif such as the ribosomal transcription factor UBF [3], the lymphoid enhancer-binding factor LEF-1 [4],

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_10,

or the yeast mitochondrial factor Abf2 [5]. These protein-induced topological changes can be readily visualized as changes in the degree of supercoiling in a closed circular DNA molecule, and the analysis thereof can be performed on a simple agarose gel.

The topology of a circular DNA molecule is defined by three parameters: the linking number *Lk* (which is the number of times one strand of the double helix goes over the other one), the writhe *W* (the path of the DNA duplex through space), and the twist *Tw* (the number of base pairs found in one turn of the DNA helix). The value of *Tw* for a relaxed B-form DNA helix equals roughly 10.5 base pairs per turn [6]. Changing this value would normally result in a topological stress in a closed, circular DNA molecule and would have to be compensated by the other parameters. A change in twist can be introduced by the presence of intercalating agents such as ethidium bromide or chloroquine. A change in the linking number, meanwhile, can only occur after the covalent break and repair of at least one of the DNA strands that would allow it to pass over the other strand in one direction or the other; topoisomerases, for example, help relax DNA by performing such an action. The writhe, finally, is obviously influenced by the other two values but can be directly acted upon by a DNA-bending activity. Changes in the three parameters *Lk*, *Wr*, and *Tw* influence each other according to the equation

$$\Delta Lk = \Delta Wr + \Delta Tw.$$

As stated above, a topological stress can be compensated in a circular molecule by the generation of supercoils or loops of the double helix upon itself (another way of describing a change in writhe). Therefore, for a constant value of *Tw*, a circular DNA molecule on which a topoisomerase activity changes the linking number by one would compensate by contorting itself into the simplest form of supercoil (Fig. 1). As the linking number can be changed in either orientation, the supercoil generated will have a value of −1 if the linking number is reduced or 1 if the linking number is increased.

In the cell, topoisomerases tend to remove supercoils since they relax topological stresses. It is the interaction of DNA with other proteins, such as occurs in the nucleosome, that can stabilize certain levels of supercoiling. In this particular case, it is generally accepted that a typical nucleosome stabilizes one negative supercoil [7].

Fig. 1 (continued) make them run faster (on the *left* of the figure). Note that the nicked plasmid running at the same position as relaxed DNA is unaffected by chloroquine. (**c**) Expected distribution of the topoisomers seen in (**a**) if the DNA is initially positively supercoiled and if the agarose gel contains chloroquine. The chloroquine here would increase the positive supercoiling of the DNA and cause it to run faster than in the conditions seen in (**a**). Comparing the migration of a plasmid in the absence and the presence of chloroquine, it become possible to ascertain the direction of its supercoiling

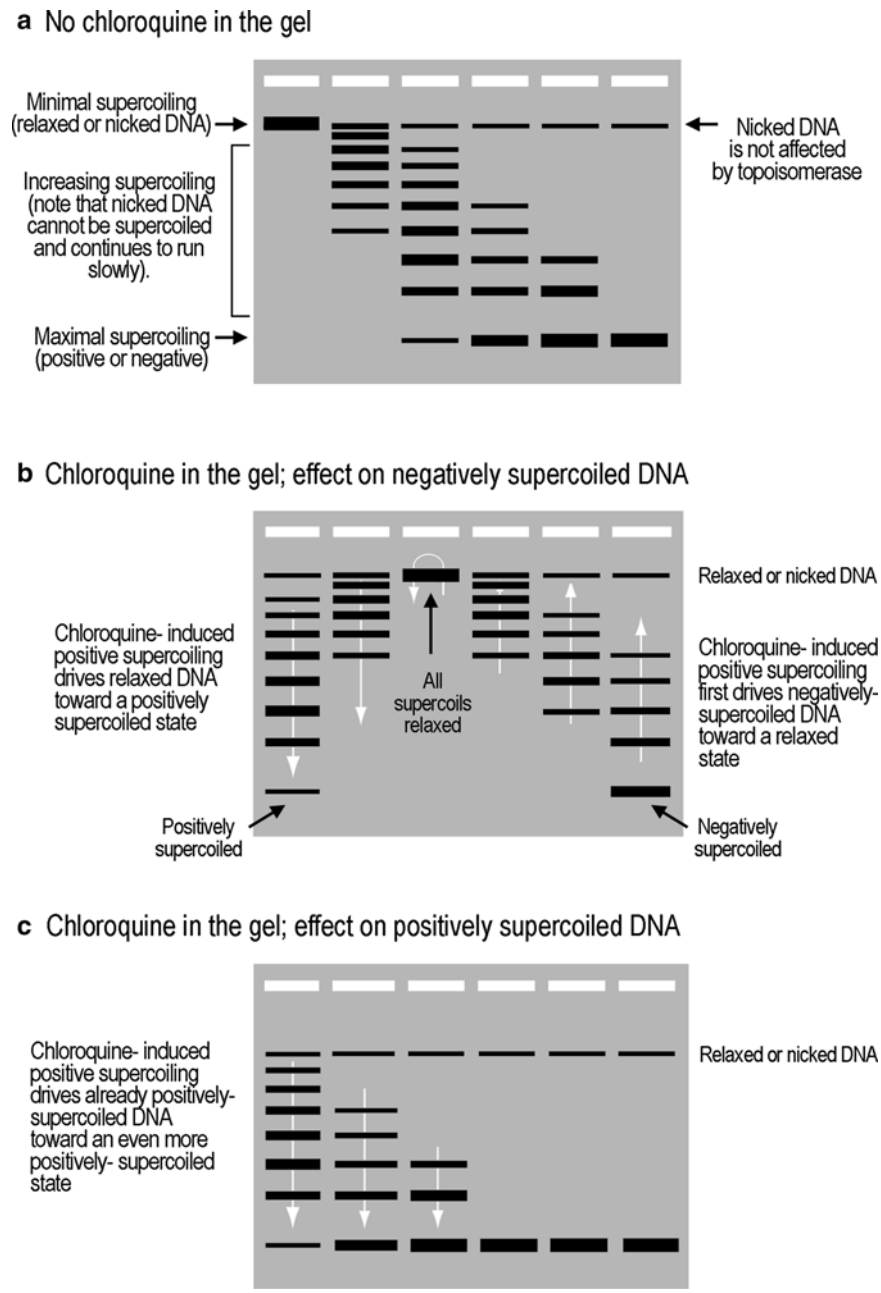


Fig. 1 Distribution of topoisomers on an agarose gel in the absence or presence of chloroquine. (**a**) Expected distribution of topoisomers with increasing levels of supercoiling on an agarose gel. Migration speed will be proportional to the degree of supercoiling, with relaxed plasmids running more slowly and progressively more supercoiled plasmids running faster and faster. Both positively and negatively supercoiled plasmids run faster than their relaxed form. Note that in all preparations of topoisomers, a small fraction of nicked plasmid is usually present and runs as a slow band. (**b**) Expected distribution of the topoisomers seen in (**a**), if the DNA is initially negatively supercoiled and if the agarose gel contains chloroquine. The effect of chloroquine is to induce positive supercoiling, which here would relax the most negatively supercoiled molecules and cause them to migrate more slowly (on the *right* of the figure) or add positive supercoils to relaxed molecules and

Circular DNA molecules like plasmids and episomes can be isolated and purified in such a way as to avoid the nicking of their phosphate backbone (which would result in their complete relaxation), and their level of supercoiling can be determined by electrophoresis on an agarose gel alongside topoisomer standards. This can be useful in determining the number of nucleosomes found on an episome, since that number would correspond to that of negative supercoils in the circular molecule [7].

Supercoiling induced by a DNA-bending protein can also be studied in vitro. To that end, the protein is incubated with a circular DNA molecule. By forcing the DNA to bend, the protein changes its writhe. Topoisomerase I is then added to the DNA–protein complex and makes it possible for the circular DNA molecule to change its linking number value, helping to dissipate the topological stress caused by the initial bending of the helix. Unlike the writhe, which can be changed with no covalent bond being broken, the linking number remains the same unless a nicking activity is present. When the DNA-binding protein is removed during the next step, along with the topoisomerase, the bending-induced change in writhe will be lost due to the removal of the DNA-binding protein, but the modified linking number will remain modified for lack of topoisomerase.

As this remaining change in linking number itself causes a topological stress in a closed circular molecule, it is in turn compensated by a change in writhe and the generation of supercoils. These supercoils, giving rise to a certain number of topoisomers, can be separated and observed on an agarose gel. Because topoisomers on a gel form distinct bands each differing from their immediate neighbors by one turn of helix (or one supercoil), it is possible to count how many turns in the linking number can be added or removed by a particular DNA–protein interaction.

To determine whether a DNA–protein interaction induces positive or negative supercoiling, the topoisomers are run on two gels in parallel, one in the absence and the other in the presence of an intercalating agent such as chloroquine [8]. Chloroquine, like ethidium bromide, inserts itself between the strands of the double helix and so artificially unwinds it, reducing the value of *Tw*. This is compensated by a forced change in writhe, and so the value of *Wr* increases. In simple terms, a plasmid run in a chloroquine gel appears more positively supercoiled than it would normally be.

As seen in Fig. 1, this can have either of two effects, depending on whether the plasmid was positively or negatively supercoiled to begin with. A supercoiled DNA plasmid runs faster than its relaxed form, irrespective of the orientation of the supercoiling. Topoisomers are therefore distributed on a gel between two positions: a very fast band that corresponds to highly supercoiled DNA and a very slow band that corresponds to relaxed (or nicked) DNA. The different topoisomers that contain more than zero

supercoils and less than the maximum possible number form a ladder between these two extremes.

If the test plasmid was initially positively supercoiled, it would run as a fast band on a normal agarose gel and even faster (being made even more positively supercoiled) on a chloroquine gel. On the other hand, if a plasmid was initially negatively supercoiled, it would run as a fast band on a normal agarose gel but as a slower band on a chloroquine gel (being made less negatively supercoiled or, in its particular case, less supercoiled, period). Note that with increasing concentrations of chloroquine, it would be possible to induce so much positive supercoiling in the latter plasmid that it would run as a slower and slower band until it reached the level of the relaxed band, after which it would start running faster and faster as an increasingly positively supercoiled molecule.

Topoisomer standards can be prepared from the test plasmid, without the use of a DNA-binding protein and using only ethidium bromide and topoisomerase I [7]. Such standards, run alongside the topoisomers generated during an experiment, can make it easier to determine how many supercoils have been introduced on average in a plasmid by a specific interaction with a protein.

2 Materials

2.1 For the DNA–Protein Interaction Reaction

1. 2× reaction buffer (20 mM Hepes pH 7,9, 80 mM KCl, 6 mM $MgCl_2$, 0.2 mM EDTA, 1 mM DTT) (*see* **Note 1**).
2. Topoisomerase I (New England Biolabs, M0301S).
3. 3 M sodium acetate.
4. TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8.0) (1 L).
5. tRNA 10 mg/mL.
6. Phenol:chloroform:isoamyl alcohol (25:24:1).
7. Ethanol.
8. Ethanol 70 %.
9. Bath at 37 °C.
10. A suitable plasmid diluted in H_2O to a working concentration of 20 ng/μL (*see* **Note 18**).
11. The protein extract to be tested.

2.2 For the Preparation of Topoisomer Standards

1. Ethidium bromide (solid).
2. Topoisomerase I (New England Biolabs, M0301S).
3. 3 M sodium acetate.
4. 10 % lauryl sulfate (SDS).
5. Phenol:chloroform:isoamyl alcohol (25:24:1).
6. Isopropanol.

7. Ethanol 70 %.
8. Isoamyl alcohol.
9. TE buffer (10 mM Tris–HCl, 1 mM EDTA, pH 8).
10. Dialysis buffer (10 mM Tris–HCl, 1 mM EDTA, 3 M NaCl, pH 8) (2 L).
11. Dialysis tubing.
12. Bath at 37 °C.
13. Stir plate (in a cold room or a refrigerated unit).

2.3 For Electrophoresis, Transfer, and Hybridization

1. Electrophoresis apparatus.
2. TBE buffer (89 mM Tris base, 89 mM boric acid, 2 mM EDTA).
3. Electrophoresis grade agarose.
4. 5× loading buffer (0.1 % bromophenol blue, 0.1 % xylene cyanol, 15 % glycerol, in TE).
5. Chloroquine diphosphate (Sigma #C6628-25G).
6. Ethidium bromide staining solution 5 μg/mL (*see* **Note 2**).
7. Southern blot apparatus.
8. Hybond N+ nylon transfer membrane or the equivalent.
9. Denaturing solution (0.5 M NaOH, 1.5 M NaCl).
10. Neutralizing solution (1 M Tris–HCl pH 8.0, 1.5 M NaCl).
11. Hybridization rotating oven and hybridization bottles.
12. 20× SSC (3 M NaCl, 0.3 M sodium citrate, pH 7.0).
13. 100× Denhardt's solution (2 % Ficoll 400, 2 % polyvinylpyrrolidone, 2 % bovine serum albumin).
14. Pre-hybridization solution (6× SSC, 10× Denhardt's solution, 0.1 % SDS, 0.1 % sodium pyrophosphate) (*see* **Note 3**).
15. Single-stranded salmon sperm DNA (ssDNA) 10 mg/mL.
16. Labeled probe (*see* **Note 4**).
17. Phosphorimager cassette (*see* **Note 5**).

3 Methods

The experiment should be performed in duplicate. One set of samples will be loaded on an agarose gel containing no chloroquine, the other set on a gel containing it (*see* **Note 13**).

3.1 Analysis of Protein-Induced DNA Bending

1. Clearly label as many Eppendorf tubes as will be required for the experiments. To each, add the appropriate amount of 2× buffer and H_2O with 20 ng of a suitable plasmid roughly 3000 base pairs long. Try to work in as small a volume as possible (*see* **Note 6**).

2. Add the protein sample to the reaction tubes. So as not to miss the most appropriate DNA–protein ratio, protein quantities should be distributed along a reasonable range and in regular increments (e.g., 0.3, 9, 27, 91, 273 ng, where each tube contains three times the amount of the previous one). Keep on ice.
3. Allow the DNA–protein interaction to proceed for 20 min (*see* **Note 7**).
4. Add 1 U of topoisomerase I and continue the incubation at 37 °C for 10 min (*see* **Note 8**).
5. To each tube, add 200 µL TE, 1 µL tRNA 10 mg/mL as carrier and adjust to 0.3 M sodium acetate (*see* **Note 9**).
6. Extract DNA with 200 µL phenol:chloroform:isoamyl alcohol; vortex briefly and centrifuge for 5 min on a table-top microcentrifuge. Transfer the aqueous (top) phase in a fresh tube.
7. Precipitate DNA with one volume isopropanol. Centrifuge as above for 5 min.
8. Wash the DNA pellet with 100 µL ethanol 70 %; centrifuge as above for 5 min. Air-dry the pellet.
9. Add 8 µL TE buffer and 2 µL 5× loading buffer. Load on a 0,7 % agarose/1× TBE gel containing either 0 or 10 µg/mL chloroquine. Run the gel at 70 V with buffer recirculation for 5–6 h or until the xylene cyanol band (light blue) has traveled 7 cm on the gel (*see* **Notes 10** and **11**). An overnight run at a lower voltage is also appropriate (*see* **Note 13**).
10. Denature the DNA by soaking the gel in the denaturing solution with gentle rocking for 30 min.
11. Neutralize the gel by soaking the gel in two changes of neutralizing solution, with gentle rocking, for 30 min each time.
12. Transfer the DNA to a Hybond N membrane using Southern blotting. Clearly mark the side on which the DNA is found.
13. Pre-hybridization: Put the membrane in a hybridization bottle, DNA side facing the inside of the bottle. Add a small volume of pre-warmed (65 °C) pre-hybridization solution as it will allow complete covering of the membrane (usually 5 mL for small bottles and 10 mL for long bottles). Denature the ssDNA by heating at 95 °C for 5 min and add to the pre-hybridization solution to a final concentration of 100 µg/mL. Allow pre-hybridization to proceed at 65 °C for 1 h, with rotation.
14. Denature the labeled probe by heating at 95 °C for 5 min and add to the pre-hybridization solution (do not pipet the undiluted probe straight on the membrane). Allow the hybridization to proceed for more than 6 h (overnight is fine) at 65 °C, with rotation.

15. Wash the membrane with multiple changes of 50 mL SSC of decreasing concentration, 30 min at a time until the desired stringency is reached (e.g., 3× SSC, 1× SSC, 0,1× SSC) (*see* **Note 12**).
16. Dry the membrane and expose to a phosphorimager cassette or to an X-ray film.

3.2 Preparation of Topoisomer Standards

It is not necessary to use topoisomer standards to determine if a particular protein induces supercoils in a plasmid or to determine the direction of the supercoiling. Such standards can however be useful tools to count how many supercoils have been introduced in a molecule, as described in ref. 7. The standards should be prepared with the same plasmid that will be used in the experiments:

1. Prepare a fresh ethidium bromide solution (2 mg/mL) from the solid stock (*see* **Note 14**). Protect from light.
2. From that solution, prepare a fresh dilution at 0.1 mg/mL. Protect from light.
3. Determine the concentration of the plasmid to be used with a spectrophotometer. Use an intact supercoiled plasmid, with as little nicked material as possible.
4. For each standard, in an Eppendorf tube, add 5 μg plasmid, 2 μL New England Biolabs buffer NEB4 (provided with the topoisomerase), 1 μL (5 U) topoisomerase I, (*x*) μL ethidium bromide 0.1 mg/mL, and (*y*) μL H_2O (where *x* and *y* could be 0 and 7.5, 1.6 and 5.9, 3.1 and 4.6, and 6.2 and 1.3, respectively, for a final volume of 20 μL). The higher the concentration of ethidium bromide, the more positively supercoiled each standard will be. Protect from light (*see* **Note 15**).
5. Incubate in the dark at 37 °C for 3 h.
6. Add 264 μL TE, 6 μL SDS 10 %, and 30 μL 3 M sodium acetate to a final volume of 300 μL (for a final concentration of 0.2 % SDS and 0.3 M sodium acetate).
7. Extract most of the ethidium bromide by adding one volume isoamyl alcohol to each tube, shaking briefly and discarding the top phase. Work in dim light (*see* **Note 16**). Repeat four times.
8. Make sure the volume has remained at roughly 300 μL; if not, top up with H_2O (*see* **Note 17**).
9. Extract the DNA with one volume of phenol:chloroform:isoamyl alcohol. Vortex and centrifuge for 3 min in a table-top microcentrifuge. Recover the top (aqueous) phase.
10. Extract the aqueous phase again with one volume chloroform. Vortex and spin as above. Recover the top phase.
11. Dialyze each standard in an independent piece of dialysis tubing. Dialysis should be performed in the cold room and in the dark. The first dialysis is against 1 L dialysis buffer, overnight.

The tubes are then dialyzed against 1 L of fresh dialysis buffer, still in the dark and in the cold room, for the rest of the day (at least 6 h). The tubes are finally dialyzed against 1 L TE buffer, still in the dark and in the cold room, for another overnight period.

12. The standards are recovered and kept in Eppendorf tubes. They should be analyzed by spectrometry or on gel to ascertain their concentration.

3.3 Analysis of the Chloroquine Gels

Topoisomers separated by gel electrophoresis are distributed between two extremes: the fastest running band, which corresponds to highly supercoiled plasmid (whether it be negatively or positively), and the slowest running band, which corresponds to relaxed plasmid (whether nicked or relaxed by topoisomerase). Plasmid DNA extracted from bacteria, as a rule, is negatively supercoiled and runs faster than its linear form.

Protein-induced supercoiling in a plasmid is apparent after the migration pattern of the generated topoisomers has been visualized and compared to the migration of the untreated plasmid.

On a normal agarose gel containing no chloroquine, any activity that introduces *positive* supercoils in a naturally negatively supercoiled plasmid would cause it to run more slowly (higher on the gel), because in effect it relaxes it somewhat. On the other hand, in the case of protein-induced *negative* supercoiling, the band might be expected to actually run faster than the untreated control since more negative supercoiling is added to the already negative supercoils. That, however, may not be visible since the control DNA might already contain as many negative supercoils as it will tolerate.

To determine if a DNA molecule has been positively or negatively supercoiled, it has to be run on two gels in parallel: one without and one with chloroquine (which introduces positive supercoils in DNA). Figure 1 shows how chloroquine will alter the course of supercoiled DNA. Panel A shows a normal gel on which a plasmid has been increasingly supercoiled by an increasing quantity of protein. Panel B illustrates what would happen if the samples seen in A were negatively supercoiled and run on a chloroquine gel. Because of the positive-inducing effect of chloroquine, the more heavily negatively supercoiled samples (on the right of the panel) are turned more positive and so more relaxed, migrating more slowly. At the same time, the samples that were not supercoiled at all (at the left of the panel) acquire positive supercoils, which drive them toward the bottom of the gel. Note that in between these extremes, there is a degree of negative supercoiling that will be exactly compensated by the positive effect of the chloroquine, resulting in a relaxed band (third track from the left). Panel C illustrates what would happen if the topoisomers seen in A were initially positively supercoiled. In such a case, since there is no negative supercoil to be compensated, all bands are driven toward the bottom of the gel by the chloroquine.

4 Notes

1. This buffer has been successfully used with the transcription factor UBF from *Xenopus laevis* on a plasmid containing repeats of the transcriptional enhancers found in the animal's ribosomal gene spacer. It should be optimized for the protein of interest, but is expected to be the same as what would be used for any DNA-binding experiment like electromobility shift assays or DNAse I footprinting. Care should also be given to that buffer's compatibility with the topoisomerase used; certain commercial enzymes have specific salt requirements and incompatibilities. Depending on the degree of purity of the protein, a nonspecific competitor such as poly(dI-dC)·poly(dI-dC) can be included in the reaction.
2. The gels must be run in the absence of any staining agent, so as not to expose the DNA to intercalating agents other than chloroquine when appropriate. They must therefore be stained after migration. Here ethidium bromide is suggested, but other stains are quite acceptable (e.g., SYBR green).
3. For 1 L, add 300 mL 20× SSC to 500 mL H_2O, then add 100 mL 100× Denhardt's solution, and only then add 5 mL SDS 20 % and 1 g pyrophosphate. Top up with H_2O. This order of addition will prevent precipitation.
4. The probe can be radioactive or fluorescent, depending on availability and preference.
5. The use of a phosphorimager system is not essential; an X-ray film-based system is perfectly suitable.
6. The DNA–protein interaction will be favored by having all reagents in as concentrated a state as possible. The same holds true for the topoisomerase activity. A reaction volume in the 20–50 μL range is advised, if the protein concentration allows it. The initial reaction volume should allow for the later addition of topoisomerase.
7. Although the reaction described here is performed on ice, some DNA–protein reactions are routinely better observed at 30 or 37 °C. The experimenter must determine the best course for any specific protein analyzed.
8. This amount of topoisomerase I, according to the supplier (New England Biolabs), should relax most of 500 ng of supercoiled plasmid in 15 min. It is however very salt sensitive, and its efficiency may vary depending on the exact composition of the reaction buffer used. Since the relaxation of the plasmid is essential to the experiment, it should be determined beforehand how much enzyme must be added to the test plasmid in order to relax it completely after 10 min at 37 °C and in the experimental buffer (not the supplier's buffer).

9. The following steps describe the purification and precipitation of the DNA. Alternatively, loading buffer can be directly added to the reaction tubes and the DNA loaded straight on the agarose gels. Although less clean because it does not guarantee the release of the protein from the DNA, this approach has worked in the past (e.g., ref. 3).
10. This represents a fairly long run, but is necessary for the topoisomers to clearly separate from one another. The faster the gel runs, the more it will heat—which could alter the gel and make the bands less sharp. Recirculation of the buffer is in such a case particularly important. Running the gel in the cold room is also an option to limit heating.
11. If a molecular weight marker is used, make sure that it is either diluted enough so as not to give too strong signal if the probe hybridizes to it nonspecifically or leave a few blank wells between it and the samples.
12. Stringency will have to be determined empirically.
13. The protocol described uses a small amount of DNA (20 ng) and requires blotting to reveal the supercoils. It would be theoretically possible to use high enough concentrations of DNA, protein, and topoisomerase to make direct visualization on the gel possible (after a post-run staining of the material), but this would require a lot of purified protein and more enzyme than is financially advisable.
14. Although ethidium bromide is not as toxic as popular laboratory folklore would have, it remains a dangerous product and should be handled with care, especially in its solid form. Avoid contact with the skin and avoid generating (and especially breathing) any ethidium bromide dust. Wear protective clothing.
15. Ethidium bromide intercalated in DNA will cause nicking when exposed to light. Although it is not necessary to work in the dark, it is better to dim the lights as much as possible and work in a shaded place (such as a chemical hood).
16. The isoamyl alcohol in the top phase will remove most of the ethidium bromide. DNA will remain in the bottom phase. As isoamyl alcohol is rather noisome and will contain ethidium bromide, it is better to work in a chemical hood (which also helps with the low light).
17. The volume is not absolutely crucial here, but to maintain it around 300 µL will help with the upcoming phenol extractions. As a crude measuring device, we use a home-made graduated Eppendorf tube where we marked volumes at 100 µL increments.
18. The plasmid should be roughly 3000 base pairs long or shorter, as longer sizes require longer separation runs on agarose.

The plasmid should contain one or multiple copies of the protein's binding site, as multiple binding will amplify any supercoiling effect caused by the DNA–protein interaction.

References

1. Simpson RT, Thoma F, Brubaker JM (1985) Chromatin reconstituted from tandemly repeated cloned DNA fragments and core histones: a model system for study of higher order structure. Cell 42:799–808
2. Javaherian K, Liu JF, Wang JC (1978) Nonhistone proteins HMG1 and HMG2 change the DNA helical structure. Science 199:1345–1346
3. Bazett-Jones DP, Leblanc B, Herfort M, Moss T (1994) Short-range DNA looping by the Xenopus HMG-box transcription factor, xUBF. Science 264:1134–1137
4. Giese K, Pagel J, Grosschedl R (1997) Functional analysis of DNA bending and unwinding by the high mobility group domain of LEF-1. Proc Natl Acad Sci U S A 94:12845–12850
5. Diffley JF, Stillman B (1992) DNA binding properties of an HMG1-related protein from yeast mitochondria. J Biol Chem 267: 3368–3374
6. Crick FH (1976) Linking numbers and nucleosomes. Proc Natl Acad Sci U S A 73: 2639–2643
7. Clark DJ (1998) Counting nucleosome cores on circular DNA using topoisomerase I. Chromatin: a practical approach. In: Gould H (ed.) Oxford University Press, Oxford
8. Shure M, Pulleyblank DE, Vinograd J (1977) The problems of eukaryotic and prokaryotic DNA packaging and in vivo conformation posed by superhelix density heterogeneity. Nucleic Acids Res 4:1183–1205

Chapter 11

Precise Identification of DNA-Binding Proteins Genomic Location by Exonuclease Coupled Chromatin Immunoprecipitation (ChIP-exo)

Dominick Matteau and Sébastien Rodrigue

Abstract

DNA-binding proteins play a crucial role in all living organisms by interacting with various DNA sequences across the genome. While several methods have been used to study the interaction between DNA and proteins in vitro, chromatin immunoprecipitation followed by sequencing (ChIP-seq) has become the standard technique for identifying the genome-wide location of DNA-binding proteins in vivo. However, the resolution of standard ChIP-seq methodology is limited by the DNA fragmentation process and presence of contaminating DNA. A significant improvement of the ChIP-seq technique results from the addition of an exonuclease treatment during the immunoprecipitation step (ChIP-exo) that lowers background noise and more importantly increases the identification of binding sites to a level near to single-base resolution by effectively footprinting DNA-bound proteins. By doing so, ChIP-exo offers new opportunities for a better characterization of the complex and fascinating architecture that resides in DNA-proteins interactions and provides new insights for the comprehension of important molecular mechanisms.

Key words Chromatin immunoprecipitation, ChIP-exo, Protein-DNA interaction, DNA-binding protein, Next-generation sequencing, DNA footprinting, Formaldehyde cross-linking, Genomics, Exonuclease degradation, Gene regulation

1 Introduction

DNA-binding proteins are essential for the regulation of many major cellular processes such as transcription, genome replication, and DNA repair. These proteins include transcription factors that regulate gene expression such as activators and repressors and histone proteins that form the core of nucleosomes to pack eukaryotic genomes into chromatin, as well as CRISPR/Cas systems used by various bacteria and archaea to mediate defense against foreign nucleic acid. Mobile genetic elements, like integrative and conjugative elements (ICEs), also express DNA-binding proteins such as transcriptional activators and recombinases to disseminate through a wide spectrum of hosts [1]. Since almost all biological functions

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_11,

ultimately depend on the recognition of specific DNA sequences throughout an entire genome [2], the identification of genomic targets recognized by DNA-binding proteins has always been an area of intensive research. This endeavor is not only crucial to extend our understanding of the basic mechanisms of the living world but is also intimately tied to some of the most important health-care issues of nowadays, such as cancer, diabetes, or dissemination of antibiotic resistance.

A large number of techniques, such as DNA electrophoretic mobility shift assay (EMSA), DNA pull-down assay, reporter gene assay, and chromatin immunoprecipitation (ChIP), have been developed to study the complex interaction between DNA and proteins. Chromatin immunoprecipitation coupled with microarrays (ChIP-on-chip) has been a groundbreaking tool for studying DNA-protein interactions in the early 2000s, mostly because this methodology enabled an in vivo genome-wide screening of DNA-binding proteins targets [3, 4]. However, ChIP-on-chip can be expensive for studying mammalian genomes due to the large amount of probes required and often suffer of high background noise caused by the hybridization procedure. With the advent of next-generation sequencing, new genomics technologies emerged and led to significant improvements in the study of DNA-protein interactions. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) has been an early application of next-generation sequencing and has rapidly become a very popular and trusted technique to identify the genome-wide location of DNA-binding proteins [5]. Instead of being hybridized on a DNA microarray, immunoprecipitated DNA fragments are directly sequenced, thus providing greater genome coverage, larger dynamic range, and fewer artifacts than ChIP-on-chip [6, 7]. ChIP-seq, as well as other high-throughput techniques like DNase I hypersensitive sites sequencing (DNase-seq) and formaldehyde-assisted identification of regulatory elements (FAIRE-seq), has been extensively used in eukaryotes to study genome-wide DNA-protein interactions, such as histone modifications and variants, nucleosome-free regions, and transcriptional regulation of gene expression [5–10]. Recently, ChIP-seq has been modified to allow the genome-wide identification of DNA-binding proteins location with an unprecedented near single-base-pair accuracy [11]. This enhanced method, called chromatin immunoprecipitation with exonuclease trimming followed by high-throughput sequencing (ChIP-exo), harnesses the power of $5' \rightarrow 3'$ exonucleases to trim immunoprecipitated DNA molecules on one strand, thereby significantly increasing the resolution and sensitivity of the original ChIP-seq methodology. The ChIP-exo method was originally published by Rhee and Pugh [12] to identify the genome-wide binding sites of yeast transcription factors Reb1, Gal4, Phd1, Rap1, and human CTCF. It was also used by the same laboratory to study the binding location of TBP, TFIIB, and Pol II, allowing the

mapping of approximately 160,000 transcription initiation complexes across the human genome [13]. Soon after, the ChIP-exo approach was also adopted by other laboratories to provide new insights into FoxA1-DNA-binding properties [14] and to study antibiotic resistance genes dissemination by characterizing the regulation of the master activator of IncA/C conjugative plasmids (AcaCD) [1].

The complete ChIP-exo methodology described in this chapter is performed over several days and contains many crucial steps as well as optional steps (*see* Fig. 1). First, cells are subjected to cross-linking with formaldehyde to covalently fix DNA-binding proteins to their corresponding DNA. Treated cells are then lysed by sonication, which also releases and shears chromatin. Proteins of interest are then immunoprecipitated along with their associated DNA sequences using a specific antibody. During immunoprecipitation, DNA is treated with $5' \rightarrow 3'$ exonucleases to trim DNA molecules on one strand, thus surpassing sonication resolution limitations. The whole cell extract (WCE) recovered from the sonication step can also be treated with exonucleases to perform an exonuclease-treated whole cell extract (WCE-exo), a technique similar to DNase-seq where all DNA molecules protected from exonuclease degradation by bound proteins are sequenced (*see* Figs. 1 and 4d). Cross-links of immunoprecipitated DNA-protein complexes are then reversed, and contaminating RNA as well as released proteins is degraded by enzymatic reactions. DNA is purified via column purification and is next used to prepare an Illumina library to sequence the protected borders of exonuclease-treated protein-DNA complexes (*see* Fig. 2). Before sequencing, the library should be validated using a Bioanalyzer instrument (*see* Fig. 3) and can also be analyzed by quantitative real-time PCR (qPCR). Upon sequencing, the resulting reads are analyzed using bioinformatics approaches, and DNA sequences bound by the targeted proteins can be precisely identified throughout the genome (*see* Fig. 4).

2 Materials

2.1 Bacterial Culture, Cross-linking, and Chromatin Shearing

1. LB broth, sterilized by autoclave.
2. Shaking incubator.
3. Spectrophotometer.
4. 37 % (w/v) formaldehyde.
5. 2.5 M glycine.
6. 50 ml conical polypropylene tubes.
7. 4 °C pre-cooled 25/50 ml centrifuge.
8. 1.5 ml microcentrifuge tubes.

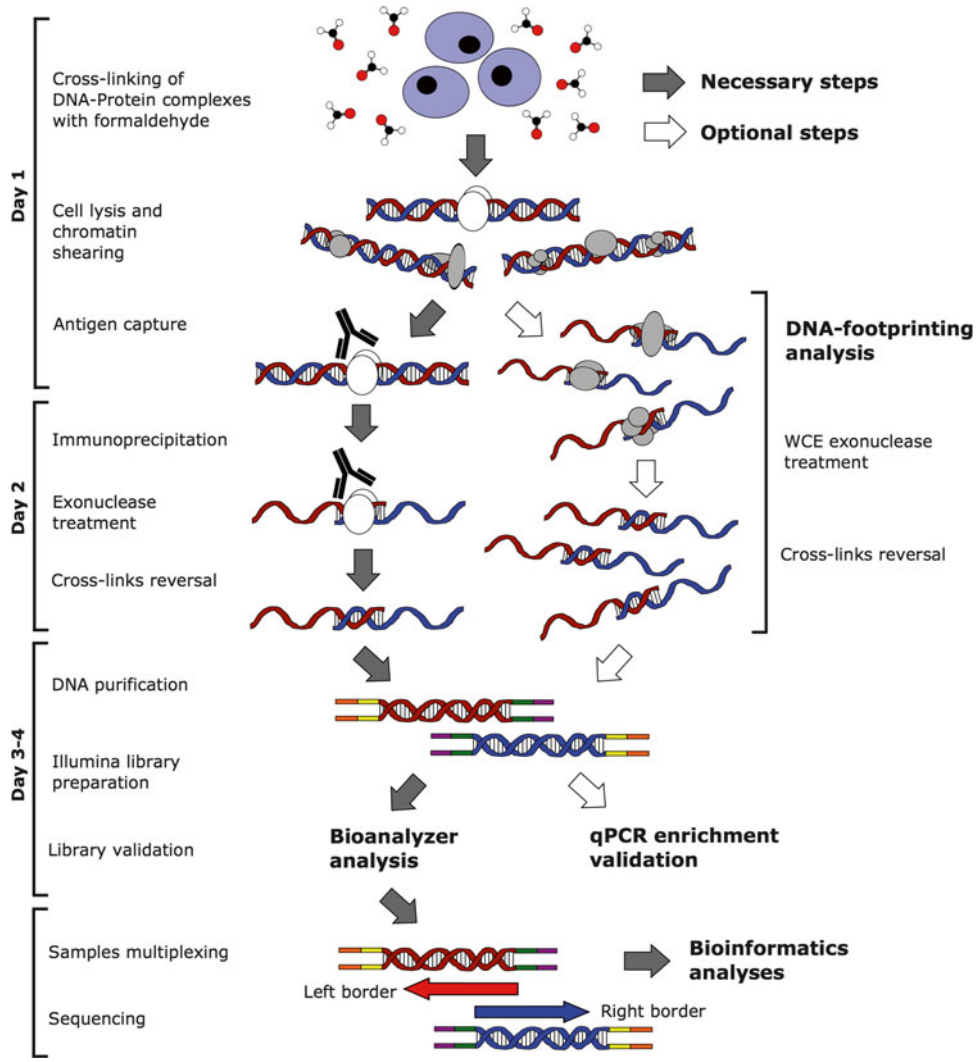


Fig. 1 Overview of the ChIP-exo methodology. Mandatory protocol steps are indicated by *gray arrows* while optional ones are shown by *white arrows*. First, cells are treated with formaldehyde to chemically cross-link DNA-protein complexes. Cells are then lysed by sonication, which also shears DNA to approximately 200–400 bp. The protein of interest (*white ellipses*) is immunoprecipitated along with its bound DNA sequences using a specific antibody. DNA sequences are then digested with $5' \rightarrow 3'$ exonucleases to trim DNA molecules up to the regions protected by DNA-bound proteins. To perform DNA footprinting analysis, the WCE obtained from the sonication step is also treated with exonucleases, but no immunoprecipitation is performed. Cross-links are then reversed, protected DNA is purified, and Illumina library is prepared. Before sequencing, the prepared library is validated by Bioanalyzer, and enrichment at specific genomic locations can be verified by qPCR

Fig. 2 (continued) polymerase at the previously digested extremities of DNA molecules (4). Nicks are repaired using the *Taq*-B DNA polymerase which exhibits a $5' \rightarrow 3'$ exonuclease activity (5). Library is finally amplified by qPCR using a high-fidelity thermophilic DNA polymerase (6, 7, and 8) with a forward oligonucleotide (IGA-PCR-PE-F, *see* Table 1) annealing on the adaptor represented by the letter *A* and a reverse oligonucleotide (TrueSeq-MPEX-R, *see* Table 1) annealing on the anchor sequence represented by the letter *B*. The reverse oligonucleotide contains an index of six nucleotides specific for each sample multiplexed for the sequencing procedure

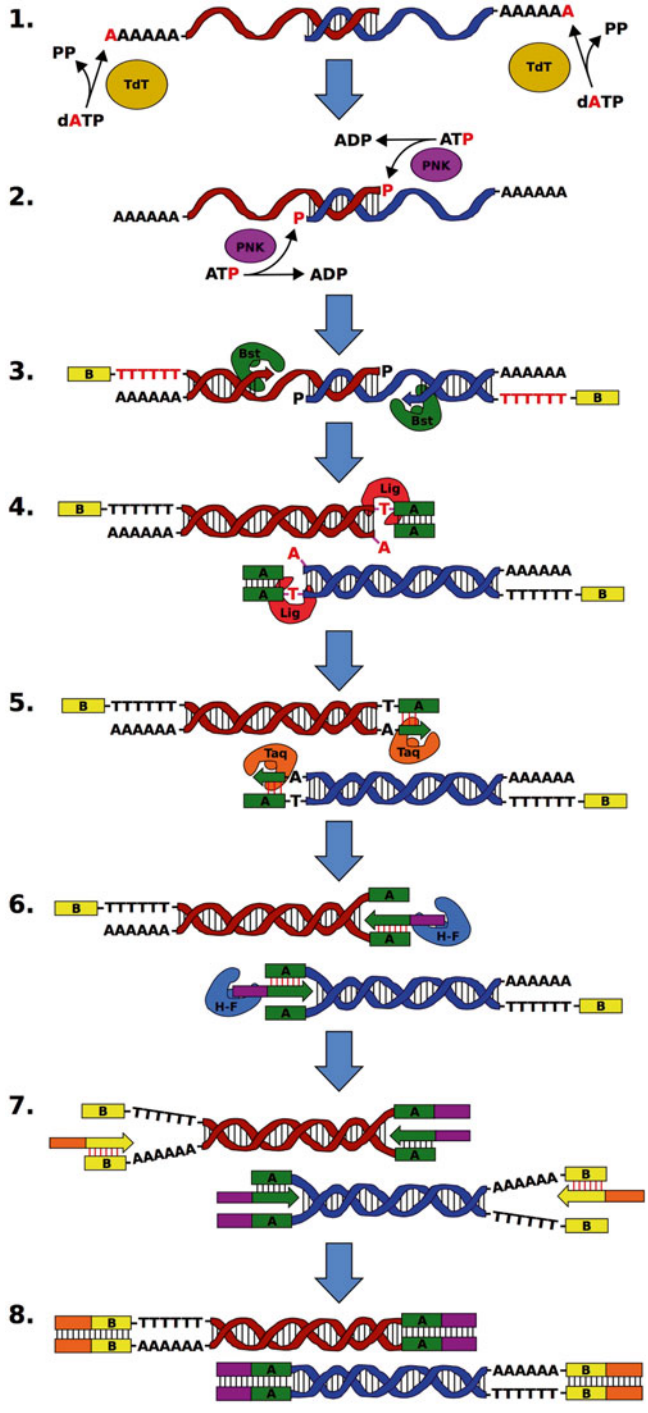


Fig. 2 ChIP-exo Illumina library preparation layout. Briefly, the 3′ ends of partially single-stranded DNA molecules are first polyadenylated (100–300 A bases) using the TdT enzyme (1), and 5′ ends are phosphorylated using T4 PNK (2). The second strand is then synthesized using the *Bst* full length DNA polymerase (3) with a poly dT oligonucleotide (TruSeq-B-dT24-VN, *see* Table 1) containing a specific anchor sequence (represented with the letter *B*) later used at the qPCR amplification step. An Illumina adaptor (represented with the letter *A*; 3′-T-IGA-A0, *see* Table 1) is next ligated using the T4 DNA ligase to the sticky adenine base left by the *Bst*

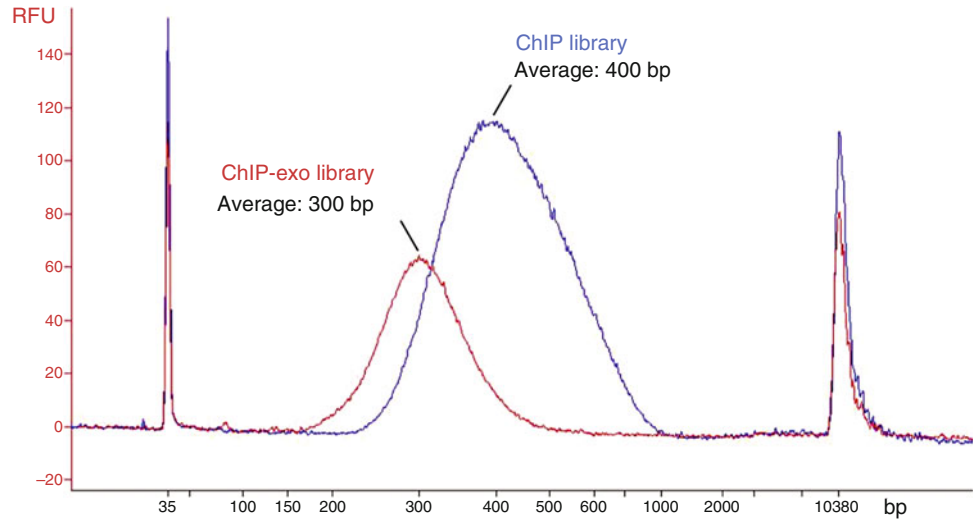


Fig. 3 Bioanalyzer profile of a ChIP-exo library. Comparison of the Bioanalyzer profile of a typical ChIP-seq library versus a typical ChIP-exo library. In general, the average size of a ChIP-exo library is slightly lower than a ChIP library due to the exonuclease treatment performed during the immunoprecipitation step (~300 bp vs. ~400 bp, respectively)

9. 4 °C pre-cooled microcentrifuge.
10. Cold TBS 1× pH 7.5.
11. 100× protease inhibitor cocktail stock solution in 90 % (v/v) ethanol: 0.2 mM pepstatin A, 72 μM leupeptin, and 26 μM aprotinin.
12. Lysis buffer : 1 % (v/v) Titron X-100, 0.1 % (v/v) Na-deoxycholate, 1 mM EDTA pH 7.5, 50 mM HEPES-KOH pH 7.5, 140 mM NaCl. Store at 4 °C and add 1 mM PMSF and 1× protease inhibitor cocktail before use.
13. Bioruptor UCD-200 sonication system (Diagenode).

2.2 Chromatin Immunoprecipitation and Exonuclease Treatment

1. Cold lysis buffer (*see* **item 12** of Subheading 2.1).
2. Antibody against the protein of interest (ideally tested for ChIP application by the manufacturer).
3. Rotating wheel.
4. Protein A or G magnetic Dynabeads (Life Technologies).
5. 1.5 ml microcentrifuge tubes.
6. Fresh BSA blocking solution: 5 mg/ml BSA, 1× PBS, store at 4 °C.
7. Magnetic support for 1.5 ml microcentrifuge tubes.
8. Cold HS lysis buffer: 1 % (v/v) Titron X-100, 0.1 % (v/v) Na-deoxycholate, 1 mM EDTA pH 7.5, 50 mM HEPES-KOH pH 7.5, 500 mM NaCl. Store at 4 °C and add 1 mM PMSF and 1× protease inhibitor cocktail before use.

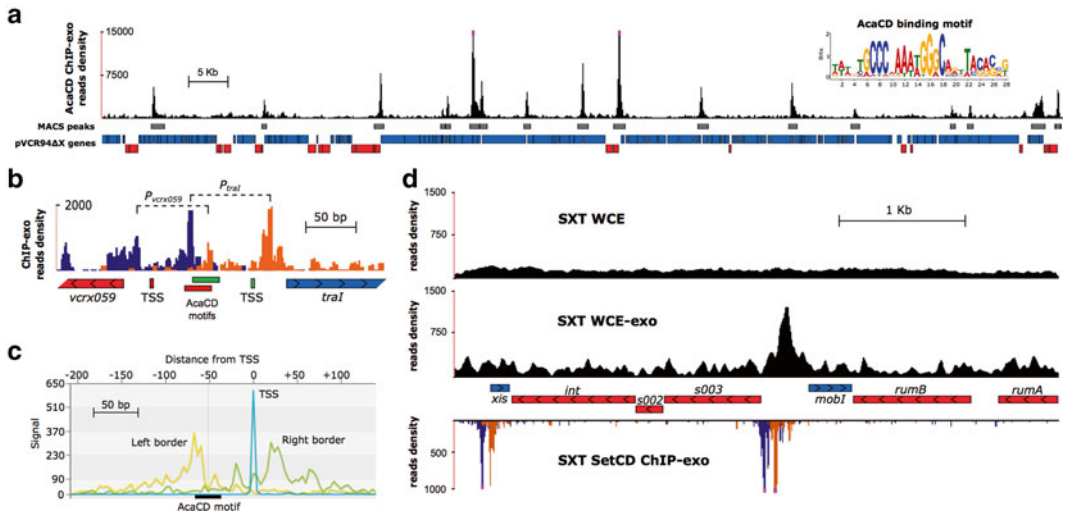


Fig. 4 Example of results generated by ChIP-exo and WCE-exo sequencing. (**a**) Results of ChIP-exo targeted on the master activator of IncA/C conjugative plasmid (AcaCD) of *Vibrio cholerae* pVCR94ΔX in an *E. coli* MG1655 background [1, 25]. The first track plots the number of ChIP-exo reads mapped on pVCR94ΔX DNA sequence (*black bars*). *Pink dots* at the *top* of peaks indicate a signal beyond the represented *y*-axis maximal value. The second track shows the position of ChIP-exo enrichment peaks found by MACS [18] (*gray*). The third track is a representation of the pVCR94ΔX genes transcribed on the positive DNA strand (*blue*) and on the negative DNA strand (*red*). The AcaCD-binding motif found by MEME [26] using ChIP-exo signals is shown at the *top right*. (**b**) Organization of pVCR94ΔX *vcrx059* and *traI* divergent promoters revealed by ChIP-exo. The first track plots ChIP-exo read density at single nucleotide resolution. *Dark blue*, density of reads mapping on the positive DNA strand; *orange*, density of reads mapping on the negative DNA strand. The second track shows the two AcaCD-binding motifs found by MAST [27] within the ChIP-exo peak between *vcrx059* (*red arrows*) and *traI* (*blue arrows*) genes. Motif corresponding to the positive DNA strand is represented in green, and motif corresponding to the negative DNA strand is shown in *red*. The exonuclease-protected regions of the *vcrx059* and *traI* promoters are indicated by *dashed lines*. (**c**) VAP aggregate profile [28] showing ChIP-exo density signals centered on the AcaCD-binding motif (*black box*). *Yellow line*, density of reads mapping on the positive DNA strand (*left border*); *green line*, density of reads mapping on the negative DNA strand (*right border*). The *X*-axis displays the distance in nucleotides from the average position of the transcription start site (TSS, *blue line*). (**d**) Example of results following sequencing of WCE and WCE-exo libraries, as well as ChIP-exo, targeted on the SetCD protein complex of SXT integrative and conjugative element (ICE) in an *E. coli* MG1655 background. The first and the second tracks plot the number of WCE and WCE-exo reads mapped on a specific region of SXT, respectively (*black bars*). The third track shows SXT genes as described in panel **a**. The fourth track plots the number of ChIP-exo reads mapped on a specific region of SXT. *Dark blue* and *orange* as described in panel **b**. *Pink dots* as described in panel **a**. While ChIP-exo reads show enrichment peaks at two distinct locations (*xis* promoter and *s003/mobI* promoter), WCE-exo reads exhibit only one enrichment peak in the promoter of *mobI*, thereby giving complementary information about the regulation of these two intergenic regions

9. Cold ChIP wash buffer: 0.5 % (v/v) NP40, 0.5 % (v/v) Na-deoxycholate, 10 mM Tris–HCl pH 8.0, 1 mM EDTA pH 8.0, 250 mM LiCl.
10. Cold TE 1×: 10 mM Tris–HCl pH 8.0, 1 mM EDTA pH 8.0.
11. Fresh ChIP-exo buffer: 1× NEBuffer 4 (New England Biolabs), 1 mM PMSF, 1× protease inhibitor cocktail.

12. T7 exonuclease (New England Biolabs, 10,000 U/ml, M0263L).
13. RecJf exonuclease (New England Biolabs, 30,000 U/ml, M0264L).
14. DNase-free RNase A (10 mg/ml).
15. Vortex mixer.
16. Multi-Therm thermomixer (Benchmark Scientific).
17. ChIP elution buffer: 1 % (w/v) SDS, 50 mM Tris–HCl pH 8.0, 10 mM EDTA pH 8.0.

2.3 Cross-links Reversal and DNA Purification

1. Cross-links reversal buffer: TE 1×, 1 % (w/v) SDS.
2. DNase-free RNase A (10 mg/ml).
3. Proteinase K (20 mg/ml).
4. ChIP DNA Clean and Concentrator Kit (Zymo Research, D5205).
5. Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, P11496).
6. Synergy HT plate reader equipped with a Take3 Multi-Volume Plate (BioTek).
7. Nanodrop.

2.4 Illumina Library Preparation

1. Thermocycler.
2. Nuclease-free 0.2 ml PCR tubes.
3. Molecular grade sterile H_2O.
4. Terminal transferase and supplied 10× buffers (New England Biolabs, 20,000 U/ml, M0315L).
5. Molecular grade 200 μM dATP.
6. Agencourt AMPure XP SPRI beads (Beckman Coulter, A63881).
7. Magnetic support for 0.2 ml PCR tubes.
8. Fresh 80 and 70 % (v/v) ethanol.
9. T4 Polynucleotide Kinase and supplied 10× reaction buffer (Enzymatics, 10,000 U/ml, Y9040L).
10. Molecular grade 10 mM ATP.
11. *Bst* DNA polymerase, Full Length, and supplied 10× reaction buffer (New England Biolabs, 5000 U/ml, M0328L).
12. Molecular grade 10 mM dNTPs.
13. TrueSeq-B-dT24-VN oligo (*see* Table 1).
14. T4 Rapid DNA ligase and supplied 2× reaction buffer (Enzymatics, 600,000 U/ml, L6030-HC-L).
15. 3′-T-IGA-A0 adaptor (*see* **Note 16** and Table 1).
16. *Taq*-B DNA polymerase and supplied 10× reaction buffer (Enzymatics, 5000 U/ml, P7250L).

Table 1
Oligonucleotides sequences used for Illumina library preparation

Name	Nucleotide sequence (5′–3′)
TruSeq-B-dT24-VN[a]	AGACGTGTGCTCTTCCGATCTTTTTTTTTTTTTTTTTTTTTTTTVN
IGA-A0-up[b,c]	/5AmMC6/ACACTCTTTCCCTACACGACGCTCTTCCGATCT
IGA-A0-down-T/A[c]	GATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAGAT
IGA-PCR-PE-F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCT
TrueSeq-MPEX-R[d]	CAAGCAGAAGACGGCATACGAGAT-INDEX-GTGACTGGAGTTCAGA CGTGTGCTCTTCCGATC

[a]Mixed bases: V:A, C, or G; N:A, T, C, or G
[b]Oligonucleotide modification /5AmMC6/: 5′ Amino Modifier C6
[c]These two oligonucleotides form the 3′-T-IGA-A0 adaptor when annealed together
[d]-INDEX- corresponds to a specific combination of 6 bp for each sample multiplexed in the library

17. VeraSeq 2.0 High-Fidelity DNA polymerase and supplied 5× reaction buffer (Enzymatics, 2000 U/ml, P7511L).
18. SYBR Green I 10×.
19. IGA-PCR-PE-F and TrueSeq-MPEX-R oligos (*see* Table 1).
20. Real-time quantitative PCR system.
21. 2100 Bioanalyzer instrument and High Sensitivity DNA Analysis Kit (Agilent Technologies, 5067-4626).

3 Methods

The current version of the ChIP-exo protocol was designed for the same purpose as the original method published by Rhee and Pugh [12] but with the following technical differences: (1) It is optimized for bacterial cells (*Escherichia coli*) instead of yeast cells and thus requires less manipulations at the cell lysis step. (2) It is optimized with magnetic Dynabeads instead of sepharose beads at the immunoprecipitation step. (3) The phenol/chloroform DNA extraction step is replaced by a more convenient column DNA purification (Zymo Research ChIP DNA Clean and Concentrator Kit) which increases DNA purity. (4) It contains fewer on-bead enzymatic reactions to prepare the sequencing library. Enzymatic reactions required for library preparation are regrouped later in the protocol. (5) Library preparation and sequencing are designed and tested using Illumina technology instead of SOLiD system but should provide equivalent results with any next-generation sequencing platform. (6) By digesting the WCE with exonucleases, it is possible to characterize the genome-wide DNA footprint of all

DNA-binding proteins (WCE-exo), a technique similar to DNase-seq.

The current described protocol was tested on a 10 ml exponential growth phase *E. coli* MG1655 culture and has to be modified accordingly if used on a different organism.

3.1 Bacterial Culture

1. Inoculate 10 ml of LB broth containing appropriate antibiotics with 100 µl of a pre-culture of the desired bacterial strain (1/100 dilution).
2. Grow cells in a shaking incubator at the appropriate temperature until the culture reaches the desired optical density (*see* **Note 1**).

3.2 Cross-Linking of DNA-Protein Complexes

See Fig. 1 for an overview of the ChIP-exo protocol.

1. Add 270 µl of 37 % formaldehyde to each 10 ml of bacterial culture (1 % final, *see* **Note 2**).
2. Incubate 20 min at room temperature with agitation (*see* **Note 3**).
3. Add 515 µl of 2.5 M glycine (0.125 M final) to quench cross-linking and incubate 5 min at room temperature with agitation.
4. Transfer bacterial culture to a 50 ml conical polypropylene tube, centrifuge the cross-linked culture 5 min at 7000×*g* (4 °C) to pellet cells, and remove the supernatant.
5. Wash cells twice with ice-cold TBS 1×, transferring cells to a 1.5 ml microcentrifuge tube after the first wash.
6. Centrifuge cells 1 min at 16,000×*g* (4 °C) and remove any trace of TBS 1× buffer (at this point, cell pellet can be snapped frozen in liquid nitrogen and stored at −80 °C).

3.3 Cells Lysis and Chromatin Shearing

1. Resuspend cells in 300 µl of cold lysis buffer and put on ice (*see* **Note 4**).
2. Sonicate samples in Bioruptor UCD-200 sonication system at HIGH amplitude for 35 cycles at 30 s ON, 30 s OFF, 4 °C (*see* **Notes 5** and **6**).
3. Centrifuge 10 min at 16,000×*g* (4 °C) to clear cell debris and transfer the supernatant to a new 1.5 ml microcentrifuge tube.
4. Optional step: Keep 5 % of the recovered volume as whole cell extract sample (WCE) and store at −20 °C (~15 µl). Proceed to optional **step 10** of Subheading 3.6 for exonuclease-treated whole cell extract (WCE-exo) sample preparation. Proceed to optional **step 7** of Subheading 3.7 for WCE sample preparation.

3.4 Antigen Capture

The following steps should be performed at 4 °C.

1. Add 500 µl of cold lysis buffer, mix thoroughly by vortexing, and keep sample on ice (800 µl final volume, *see* **Note 7**).
2. Add the desired amount of antibody to perform the antigen capture (*see* **Note 8**).
3. Incubate overnight at 4 °C on a rotating wheel.

3.5 Preparation of Protein A- or G-Coated Magnetic Beads

The following steps should be performed at 4 °C. For proper time management, it is also recommended to execute this procedure on the same day as the antigen capture (*see* Subheading 3.4).

1. Take 50 µl of protein A or G magnetic Dynabeads (i.e., 2×10^7 beads) and put in a 1.5 ml microcentrifuge tube (*see* **Note 9**).
2. Wash beads three times with 1 ml of fresh BSA blocking solution using a magnetic support for 1.5 ml microcentrifuge tubes (*see* **Note 10**).
3. After the last wash, resuspend beads in 50 µl of BSA blocking solution and incubate at 4 °C on a rotating wheel for at least 4 h.

3.6 Immunoprecipitation and Exonuclease Treatment

Steps 1–4, as well as **steps 6** and **7**, should be performed at 4 °C.

1. Transfer 50 µl of the magnetic beads solution from the previous step to 800 µl of antibody-antigen solution from **step 3** of Subheading 3.4, mix gently by pipetting, and incubate at 4 °C on a rotating wheel for at least 4 h.
2. Using a magnetic support for 1.5 ml microcentrifuge tubes, wash beads two times with 1 ml of lysis buffer, two times with 1 ml of HS lysis buffer, two times with 1 ml of ChIP wash buffer, and once with 1 ml of TE 1× (*see* **Note 10**).
3. Remove any remaining liquid and resuspend beads in 100 µl of ChIP-exo buffer for exonuclease treatment.
4. Directly add 4 µl of T7 exonuclease, 2 µl of RecJf, and 20 µg of RNase A and mix gently (*see* **Note 11**).
5. Incubate 60 min at 37 °C in a thermomixer set at 1100 rpm to avoid bead sedimentation during the enzymatic reaction.
6. Repeat bead wash (**step 2**) and remove any remaining liquid after the last wash.
7. Add 50 µl of ChIP elution buffer and vortex briefly to resuspend beads.
8. Incubate at 65 °C for 10 min and vortex every 2 min to elute DNA-protein complexes from beads.
9. Using a magnetic support for 1.5 ml microcentrifuge tube, recover supernatant and transfer it to a new tube (ChIP-exo sample).

10. Optional step: Take 1 µl of the WCE sample previously frozen at **step 4** of Subheading 3.3 and transfer in 50 µl of ChIP-exo buffer for exonuclease treatment (WCE-exo sample). Repeat **steps 4** and **5** of current subheading and proceed to the optional **step 7** of Subheading 3.7 for cross-links reversal.

3.7 Cross-Links Reversal and DNA Purification

1. Add 180 µl of cross-links reversal buffer to the 50 µl ChIP-exo sample.
2. Mix thoroughly by vortexing and incubate at 65 °C overnight to reverse cross-links.
3. Add 20 µg of RNase A directly to the tube and incubate at 37 °C for 30 min.
4. Add 11.5 µl of proteinase K directly to the mix and incubate 2 h at 37 °C.
5. Purify DNA with the ChIP DNA Clean and Concentrator Kit (Zymo Research) according to the manufacturer's specifications, except that 7 volumes of DNA-binding buffer must be used instead of 5 volumes to prevent loss of ssDNA molecules.
6. Measure DNA concentration by Quant-iT Picogreen dsDNA Assay Kit using a BioTek Synergy HT plate reader equipped with a Take3 Multi-Volume Plate (*see* **Note 12**).
7. Optional step: Respectively add 180 and 215 µl of cross-links reversal buffer to the 50 µl WCE-exo sample from optional **step 10** of Subheading 3.6 and to the 15 µl WCE sample previously frozen at optional **step 4** of Subheading 3.3. Repeat **steps 2–5** of current subheading and measure DNA concentration by nanodrop. Proceed to **step 1** of Subheading 3.8 for WCE-exo and WCE Illumina libraries preparation.

3.8 Polyadenylation and 5′ Ends Phosphorylation for Illumina Library Preparation

See Fig. 2 for an overview of the Illumina library preparation procedure. This procedure must be followed for ChIP-exo, no-Ab-IP, WCE-exo, and WCE samples before sequencing or qPCR steps.

1. Mix the following reagents in a 0.2 ml PCR tube to prepare the polyadenylation reaction (*see* **Note 13**):

   1–14.75 µl DNA (5 ng).

   2 µl 10× TdT reaction buffer.

   2 µl 10× $CoCl_2$.

   1 µl 200 µM dATP.

   0.25 µl terminal transferase.

   Complete to 20 µl with molecular grade H_2O.
2. Incubate the reaction 30 min at 37 °C in a thermocycler and then 10 min at 70 °C to inactivate enzyme.

3. Purify DNA by adding and mixing 36 μl of DNA SPRI magnetic beads (AMPure XP) to the solution (1.8:1 volume ratio, *see* **Note 14**).
4. Incubate 5 min at room temperature.
5. Place the reaction tube on a 0.2 ml magnetic support and wait 2–3 min until all beads are immobilized on the side of the tube.
6. Remove supernatant carefully while keeping the tube on the magnetic support.
7. Add 150 μl of fresh 80 % ethanol, wait approximately 30 s, and remove the liquid while keeping the tube on the magnetic support.
8. Repeat **step 7** once and remove any trace of remaining 80 % ethanol using a 1–10 μl tip.
9. Incubate the beads at room temperature for 10–15 min to dry residual ethanol (*see* **Note 15**).
10. Remove the tube from the magnetic support and resuspend beads with 16 μl of molecular grade H_2O by pipetting thoroughly but carefully.
11. Incubate at room temperature for at least 1 min.
12. Place the tube again on the magnetic support and wait 1–2 min until all beads are immobilized on the side of the tube.
13. Transfer 15 μl of the supernatant to a new 0.2 ml PCR tube (leave 1 μl of the supernatant in the tube to avoid bead contamination).
14. Mix the following reagents in a 0.2 ml PCR tube to prepare the 5′-ends phosphorylation reaction:

    15 μl polyadenylated DNA from last step.

    2 μl 10× T4 PNK reaction buffer.

    2 μl 10 mM ATP.

    1 μl T4 Polynucleotide Kinase.

    Total: 20 μl.
15. Mix well and incubate reaction at 37 °C for 30 min in a thermocycler.
16. Repeat **steps 3–12** to purify DNA with SPRI magnetic beads, except that DNA elution is done in 15 μl of molecular grade H_2O.
17. Transfer 14 μl of the supernatant to a new 0.2 ml PCR tube (leave 1 μl of the supernatant in the tube to avoid bead contamination).

3.9 Second Strand Synthesis and Adapter Ligation

1. Mix the following reagents in a 0.2 ml PCR tube to prepare the second strand synthesis reaction:

   14 μl purified DNA from last step.

2 µl 10× ThermoPol reaction buffer.

2 µl 10 mM dNTPs.

1 µl 10 µM TrueSeq-B-dT24-VN (*see* Table 1).

1 µl *Bst* DNA Polymerase, Full Length.

Total: 20 µl.

2. Mix well and incubate reaction from 40 to 70 °C for 1 h (increment of 0.5 °C/min) using a thermocycler.
3. Purify DNA by adding and mixing 20 µl of DNA SPRI magnetic beads (AMPure XP) to the solution (1:1 volume ratio, *see* **Note 14**).
4. Repeat **steps 4–12** of Subheading 3.8, except that 70 % ethanol is used instead of 80 % and that DNA elution is done in 15 µl of molecular grade H_2O.
5. Transfer 14 µl of the supernatant to a new 0.2 ml PCR tube (leave 1 µl of the supernatant in the tube to avoid bead contamination).
6. Mix the following reagents in a 0.2 ml PCR tube to prepare the adapter ligation:

   14 µl purified dsDNA from the last step.

   4 µl 100 nM 3′-T-IGA-A0 adaptor (*see* **Note 16** and Table 1).

   20 µl 2× Quick Ligation buffer.

   2 µl T4 DNA ligase.

   Total: 40 µl.
7. Mix well and incubate reaction at room temperature for 20 min.
8. Add 20 µl of molecular grade H_2O to the reaction tube to reduce PEG concentration before DNA SPRI purification.
9. Purify DNA by adding and mixing 60 µl of DNA SPRI magnetic beads (AMPure XP) to the solution (1:1 volume ratio, *see* **Note 14**).
10. Repeat **steps 4–12** of Subheading 3.8, except that 70 % ethanol is used instead of 80 % and that DNA elution is done in 19 µl of molecular grade H_2O.
11. Transfer 18 µl of the supernatant to a new 0.2 ml PCR tube (leave 1 µl of the supernatant in the tube to avoid bead contamination).

3.10 Nick Translation

1. Mix the following reagents in a 0.2 ml PCR tube to prepare the nick translation reaction:

   18 µl adapter ligated DNA from the last step.

   2.25 µl 10× *Taq*-B reaction buffer.

0.25 μl 10 mM dNTPs.

1 μl molecular grade H_2O.

1 μl *Taq*-B DNA polymerase.

Total: 22.5 μl.

2. Mix well and incubate at 66 °C for 20 min.
3. Purify DNA by adding and mixing 22.5 μl of DNA SPRI magnetic beads (AMPure XP) to the solution (1:1 volume ratio, *see* **Note 14**).
4. Repeat **steps 4–12** of Subheading 3.8, except that 70 % ethanol is used instead of 80 % and that DNA elution is done in 21 μl of molecular grade H_2O.
5. Transfer 20 μl of the supernatant to a new 0.2 ml PCR tube (leave 1 μl of the supernatant in the tube to avoid bead contamination).

3.11 Library Amplification

1. Prepare two reaction tubes (amplification duplicates) with the following reagents to amplify the ChIP-exo library by qPCR (or no-Ab-IP, WCE, and WCE-exo libraries):

   4 μl purified nick-translated DNA from the last step.

   13.2 μl molecular grade H_2O.

   5 μl 5× VeraSeq reaction buffer.

   1 μl 10 μM IGA-PCR-PE-F oligo (*see* Table 1).

   1 μl 10 μM TrueSeq-MPEX-R oligo (contains index, *see* Table 1).

   0.5 μl 10 mM dNTPs.

   0.63 μl 10× SYBR Green I.

   0.25 μl VeraSeq 2.0 DNA polymerase.

   Total: 25 μl (×2 reactions)
2. Mix well, spin down, and amplify DNA by real-time quantitative PCR using the following program:

Denature at 98 °C for 30 s	
98 °C for 15 s 60 °C for 15 s 72 °C for 15 s	Repeat amplification cycle (*n*) times.
Final extension at 72 °C for 2 min.	

3. Stop amplification reaction cycles during the late exponential phase and proceed to the final extension step (*see* **Note 17**).

4. Pool the two PCR reactions (50 µl) and purify DNA by adding and mixing 50 µl of DNA SPRI magnetic beads (AMPure XP) to the solution (1:1 volume ratio, *see* **Note 14**).
5. Repeat **steps 4–12** of Subheading 3.8, except that 70 % ethanol is used instead of 80 % and that DNA elution is done in 16 µl of molecular grade H_2O.
6. Transfer 15 µl of the supernatant to a new 0.2 ml PCR tube (leave 1 µl of the supernatant in the tube to avoid bead contamination).
7. Evaluate library size and concentration using a 2100 Bioanalyzer instrument with a High Sensitivity DNA Analysis Kit (Agilent Technologies, *see* **Note 18** and Fig. 3).
8. Optional step: Verify ChIP-exo enrichment by qPCR analysis on specific regulatory elements known to interact with the targeted protein and negative control regions. Use WCE library or no-Ab-IP library as a non-enriched sample for qPCR normalization (*see* **Note 19**).

3.12 DNA Sequencing and Data Analysis

1. If more than one sample is prepared, mix indexed libraries to obtain sufficient coverage of the corresponding reference genome(s) (*see* **Note 20**).
2. Perform single-end or paired-end sequencing on multiplexed samples at your Illumina sequencing service provider.
3. If needed, separate Illumina reads (demultiplexing) with NovoBarcode (www.novocraft.com) based on the index used for library amplification (TrueSeq-MPEX-R oligo, *see* Table 1).
4. Verify quality of raw sequence data obtained from the sequencing step using the FastQC quality control program (www.bioinformatics.babraham.ac.uk/projects/fastqc/).
5. Align forward the read of mate pair on reference genome(s) using the Burrows-Wheeler Aligner (BWA) [15] or equivalent.
6. Verify mapping quality using SAMStat [16].
7. Filter alignment using SAMtools view [17] to discard aligned reads with a quality score below 10.
8. Sort alignment using SAMtools sort [17].
9. To calculate ChIP-exo full read density as well as enrichment peaks, use the Model-based Analysis for ChIP-Seq software (MACS) [18]. Aligned reads can be converted to WIG files by MACS and then visualized with Integrative Genomics Viewer (IGV) [19] or UCSC Genome Browser [20] (*see* Fig. 4).
10. To calculate ChIP-exo 1 bp read density and to obtain near single nucleotide resolution, treat aligned reads to conserve only the first nucleotide of each read and then calculate density

with Bedtools genomecov [21]. BedGraph files output by Bedtools can also be visualized with IGV [19] or UCSC Genome Browser [20] (*see* Fig. 4).

11. Optional step: Mix the WCE and WCE-exo libraries to the ChIP-exo library and perform **steps 2–10**. Aligned reads can be used to characterize the genome-wide DNA footprint of all DNA-binding proteins and may serve as complementary data to the ChIP-exo results (*see* **Note 21** and Fig. 4).

4 Notes

1. Cells can be harvested at different optical densities depending on experimental needs. In our experience, harvesting cells during the mid-late exponential growth phase (about 0.5–0.6 of OD600 for an *E. coli* culture) yields good results.
2. Proceed rapidly to prevent cells from responding to changing environmental conditions. Formaldehyde solution may be prewarmed to avoid temperature shifts in the culture.
3. For most proteins, cross-linking 20 min at room temperature is sufficient and should give good results. Depending on many variables such as target protein abundance in the cell and its relative binding affinity to DNA, some proteins may require different cross-linking time or temperature.
4. The volume of lysis buffer used can be adapted according to the appropriate sonicator. We used Diagenode Bioruptor UCD-200 sonication system with the 1.5 ml tube holder, for which the volume of the sample must be between 100 and 300 µl.
5. Sonication settings such as intensity, time ON/OFF, and number of cycles have to be adjusted depending on the cell type used, the cellular concentration, the lysis buffer composition, and the volume of the sample.
6. The ideal size of DNA fragments after sonication is between 200 and 400 bp. Oversonication or incomplete DNA fragmentation can affect the Illumina library construction procedure and may lead to poor results. For more additional information, *see* **Note 2** and Fig. 2 of Bianco S. et al. Chapter 16 of this book.
7. At this point, sample can be split in two (400 µl each) and completed to 800 µl with cold lysis buffer to prepare a no-antibody control (no-Ab-IP). For this control, skip **step 2** of Subheading 3.4 and follow the rest of the procedure. This control can be used in qPCR analysis for enrichment validation (*see* **Note 19** for more details).

8. The concentration of antibody required for the antigen capture is normally supplied by the antibody provider but may need to be determined experimentally. In general, 2–4 μg of antibody per ChIP is sufficient.
9. The choice of the type of beads depends on the antibody affinity to protein A or G [22, 23]. For instance, mouse IgG1 antibodies have a strong affinity for protein G but show almost no affinity for protein A. Agarose or sepharose beads could also be used, provided that the bead washing procedure is modified accordingly.
10. Procedure to wash magnetic Dynabeads (all steps at 4 °C):
    (a) Place beads solution on the magnetic support for 1.5 ml microcentrifuge tubes.
    (b) Wait for about 1 min until all the beads are immobilized on the side of the tube.
    (c) Add the desired volume of the appropriate buffer and mix gently by inverting the tube a few times.
    (d) Place the tube containing the magnetic bead solution on a rotating wheel and incubate for 5 min at 4 °C.
    (e) Remove the tube from the rotating wheel and repeat the procedure if necessary.
11. Given that a 10 ml *E. coli* culture (LB broth at 37 °C) contains approximately 10^9 cells at OD600 of 0.5–0.6, that the average molecular weight of DNA is 660 g/mol/bp, and that *E. coli* contains a genome of 4.6×10^6 bp, approximately 5–10 μg of DNA has to be digested by the exonucleases. The required amount of T7 and RecJf exonucleases for DNA digestion was calculated accordingly and can be adapted if necessary.
12. Purified DNA from a ChIP-exo sample has generally a very low concentration (sometimes less than 1 ng/μl), and quantification requires sensitive methodologies such as fluorescence quantification. When measuring low DNA concentration by the Quant-iT PicoGreen dsDNA Assay Kit and a BioTek Synergy HT plate reader equipped with a Take3 Multi-Volume Plate, use the following parameters: detection method, fluorescence (excitation 485/20 and emission 528/20); read type, endpoint; optics position, top; gain, auto, scale to high wells, and scale value 60,000; delay after plate movement, 10 ms; measurements per data point, 255; delay between measurements, 1 ms; read height, 1.00 mm.
13. The concentration of dATP used is adjusted to add a 100–300 bp dA tail to 3′ ends of DNA fragments. This dA tail length is obtained by incubating the TdT with a DNA/dATP reaction ratio of 1:5000. If we assume that the molecular weight of a 200 bp fragment (average size of DNA fragments produced

after the exonuclease treatment) is approximately 132 ng/pmol (660 g/mol/bp × 200 bp), then 5 ng of DNA corresponds to ~40 fmol, which means that 200 pmol of dATP must be used in the reaction (40 fmol × 5000). If a different amount of DNA is used in the reaction, the concentration of dATP added should be modified accordingly.

14. Always use DNA SPRI magnetic beads warmed at room temperature for optimal binding efficiency. Modifying the SPRI volume ratio affects beads' capacity to bind small DNA fragments [24]. A SPRI volume ratio of 1.8:1 allows the recovery of DNA fragments larger than 80 bp. A SPRI volume ratio of 1:1 allows the recovery of DNA fragments larger than 100 bp.
15. It is important to let the beads dry completely to avoid ethanol contamination and maximize DNA elution efficiency. Generally, waiting 10–15 min at room temperature is sufficient to eliminate residual ethanol, but beads should not be over-dried since this could also lead to poor recovery.
16. The 3′-T-IGA-A0 adaptor is prepared by mixing the IGA-A0-up and IGA-A0-down-T/A (*see* Table 1) to a final concentration of 40 μM each in a 10 mM Tris-HCl and 10 mM NaCl buffer. Anneal oligos by incubating from 98 to 4 °C (1 °C/30 s decrease) in a thermocycler. For adapter ligation, use a 10× molar excess of adaptor, so with 4 μl of 100 nM adaptors (400 fmol) for approximately 5 ng of starting material (5 ng/(660 g/mol/bp × 200 bp) = 40 fmol).
17. The number of PCR amplification cycles depends on the amount of starting material. Generally, between 12 and 20 cycles is enough to reach the late exponential amplification phase. It is very important to stop the qPCR amplification reaction during the 72 °C step (extension) to keep DNA double-stranded.
18. After amplification, libraries often have a concentration between 10 and 30 ng/μl and must be diluted 1/20 for Bioanalyzer analysis using a High Sensitivity DNA assay chip. Alternatively, size distribution can be evaluated by agarose gel electrophoresis, and DNA quantification can be obtained using a nanodrop.
19. If possible, we strongly encourage verifying the enrichment of specific regulatory elements known to interact with the targeted protein by qPCR before sending the libraries for sequencing. Design primers that will target the region of interest (GOI) as well as primers that anneal to a negative control region (CTL). The WCE and the no-Ab-IP libraries can be used as non-enriched samples for qPCR normalization. The ChIP-exo enrichment can be calculated according to the ΔΔCT method as follows: $\Delta\Delta CT = (CT^{\text{no Ab/GOI}} - CT^{\text{Ab/GOI}}) - (CT^{\text{no Ab/CTL}} - CT^{\text{Ab/CTL}})$ or $(CT^{\text{WCE/GOI}} - CT^{\text{Ab/GOI}}) - (CT^{\text{WCE/CTL}} - CT^{\text{Ab/CTL}})$. Relative fold enrichment is then obtained as follows: $2^{-\Delta\Delta CT}$.

20. The coverage value used for a ChIP-seq experiment is generally around 100× but can be modified according to the desired sequencing depth and experimental needs. Sample coverage is calculated as follows:

    (sequencing reads length in bp × theoretical number of reads)/genome size in bp.

    The theoretical number of reads for one sample is calculated by multiplying the total number of reads normally obtained with the chosen sequencing technology by the sample pooling ratio (ng of sample in mix/total ng of pooled mix).

21. The WCE-exo sample can be used to identify a large number of putative protein binding site genome-wide. This method is very similar to DNase-seq [10] and can provide valuable protein-DNA interaction information to serve as complementary data to ChIP-exo results (*see* Fig. 4).

Acknowledgment

We are grateful to Nicolas Carraro and Vincent Burrus for the gift of the *E. coli* pVCR94ΔX strain. We thank Pierre-Étienne Jacques and Stéphanie Bianco for technical assistance, the *Centre de calcul scientifique* of Université de Sherbrooke for computational resources and technical support, as well as Vincent Baby for his precious comments on the manuscript. This work was supported by the *Fonds québécois de la recherche sur la nature et les technologies* through an MSc scholarship awarded to D.M. and a Projet de recherche en équipe grant awarded to S.R., and Vincent Burrus. S.R. holds a Chercheur boursier Junior 1 award from the Fonds de recherche Québec-Santé.

References

1. Carraro N, Matteau D, Luo P et al (2014) The master activator of IncA/C conjugative plasmids stimulates genomic islands and multidrug resistance dissemination. PLoS Genet 10:e1004714
2. Halford SE, Marko JF (2004) How do site-specific DNA-binding proteins find their targets? Nucleic Acids Res 32:3040–3052
3. Ren B, Robert F, Wyrick JJ et al (2000) Genome-wide location and function of DNA binding proteins. Science 290:2306–2309
4. Buck MJ, Lieb JD (2004) ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. Genomics 83:349–360
5. Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. Nat Rev Genet 13:840–852
6. Johnson DS, Mortazavi A, Myers RM et al (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316:1497–1502
7. Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet 10:669–680
8. Boyle A, Davis S, Shulha H et al (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132:311–322

9. Thurman R, Rynes E, Humbert R (2012) The accessible chromatin landscape of the human genome. Nature 489:75–82
10. Boyle AP, Song L, Lee BK et al (2011) High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. Genome Res 21:456–464
11. Rhee HS, Pugh BF (2012) ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. Current protocols in molecular biology, Unit 21.24. Wiley, New York
12. Rhee HS, Pugh BF (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell 147:1408–1419
13. Rhee HS, Pugh BF (2012) Genome-wide structure and organization of eukaryotic pre-initiation complexes. Nature 483:295–301
14. Serandour A, Brown GD, Cohen JD et al (2013) Development of an Illumina-based ChIP-exonuclease method provides insight into FoxA1-DNA binding properties. Genome Biol 14:R147
15. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760
16. Lassmann T, Hayashizaki Y, Daub CO (2011) SAMStat: monitoring biases in next generation sequencing data. Bioinformatics 27:130–131
17. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079
18. Zhang Y, Liu T, Meyer C et al (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137
19. Thorvaldsdóttir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192
20. Kent WJ, Sugnet CW, Furey TS et al (2002) The human genome browser at UCSC. Genome Res 12:996–1006
21. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842
22. Phillips T (2005) Affinity chromatography in antibody and antigen purification. Handbook of affinity chromatography, 2nd edn. CRC, Boca Raton, pp 367–397
23. Svotelis A, Gévry N, Gaudreau L (2009) Chromatin immunoprecipitation in mammalian cells. In: Moss T, Leblanc B (eds) DNA-protein interactions. Humana, New Jersey, pp 243–251
24. Rodrigue S, Materna AC, Timberlake SC et al (2010) Unlocking short read sequencing for metagenomics. PLoS One 5:e11840
25. Carraro N, Sauvé M, Matteau D et al (2014) Development of pVCR94ΔX from *Vibrio cholerae*, a prototype for studying multidrug resistant IncA/C conjugative plasmids. Front Microbiol 5:44
26. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proc Int Conf Intell Syst Mol Biol 2:28–36
27. Bailey T, Gribskov M (1998) Combining evidence using p-values: application to sequence homology searches. Bioinformatics 14:48–54
28. Coulombe C, Poitras C, Nordell-Markovits A et al (2014) VAP: a versatile aggregate profiler for efficient genome-wide data representation and discovery. Nucleic Acids Res 42:W485–W493

Chapter 12

The Cruciform DNA Mobility Shift Assay: A Tool to Study Proteins That Recognize Bent DNA

Victor Y. Stefanovsky and Tom Moss

Abstract

So-called architectural DNA-binding proteins such as those of the HMGB-box family induce DNA bending and kinking. However, these proteins often display only a weak sequence preference, making the analysis of their DNA-binding characteristics difficult if not impossible in a standard electrophoretic mobility shift assay (EMSA). In contrast, such proteins often bind prebent DNAs with high affinity and specificity. A synthetic cruciform DNA structure will often provide an ideal binding site for such proteins, allowing their affinities for both bent and linear DNAs to be directly and simply determined by a modified form of EMSA.

Key words Architectural proteins, HMGB-box protein, Bent DNA, Cruciform, EMSA

1 Introduction

The first reported interaction of an HMGB-box protein (HMGB1) with a bent stable synthetic DNA structure suggested that this family of proteins displayed an intrinsic affinity for bent DNA [1]. The assay was based on the electrophoretic mobility shift assay (EMSA) of an in vitro-assembled synthetic cruciform when it was bound by an HMGB-box. It was found that binding to the prebent cruciform DNA occurred with a much higher affinity and specificity than to linear DNA. Shortly after, the same group demonstrated an identical behavior for the sequence-specific transcription factor SRY, which contains a single HMG box [2], followed by similar reports on the other HMGB-box proteins LEF-1/TCF-1 [3, 4]. The RNA polymerase I transcription factor UBF, like HMGB1, is a non-sequence-specific DNA-binding protein with multiple HMGB-boxes, and as such its binding properties were difficult to analyze by EMSA using linear DNA. However, it was found to bind strongly to cruciform DNA [5]. A detailed study showed that the individual HMGB-boxes 1 and 2 of UBF, responsible for the rDNA promoter in-phase bending and enhancesome formation, [6, 7] each bind with high affinity to

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_12,

cruciform structures [8]. Moreover, competition with linear DNA proved to be a valuable tool for detecting changes in the binding affinity of these boxes as a result of mutations or posttranslational modifications. Such changes are believed to reflect structural changes in the protein–DNA complex due to an altered bending capacity of the HMGB-boxes.

The 14-3-3 family of proteins were found to specifically recognize cruciform structures at origins of DNA replication in a cell-cycle-dependent manner and were identified as regulators of DNA replication [9], and the cruciform mobility shift assay has also been used for analyzing these proteins.

The cruciform mobility shift assay has, thus, shown itself to be a useful tool in the study of proteins that display enhanced affinity for bent DNA. Here we present the basic assay using the HMGB-boxes as example.

2 Materials

Stock solutions of cruciform oligonucleotides at 10 pmol/μl in ddH_2O:

Oligo 1 5′-CCCTATAACCCCTGCATTGAATTCCAGTCTG ATAA-3′

Oligo 2 5′-GTAGTCGTGATAGGTGCAGGGGTTATAGGG-3′

Oligo 3 5′-AACAGTAGCTCTTATTCGAGCTCGCGCCCTAT CACGACTA-3′

Oligo 4 5′-TTTATCAGACTGGAATTCAAGCGCGAGCTCGA ATAAGAGCTACTGT-3′

1. These oligonucleotides are designed in order to anneal with each other and form a cruciform structure (Fig. 1).
2. [γ – ^{32}P]-ATP (PerkinElmer).

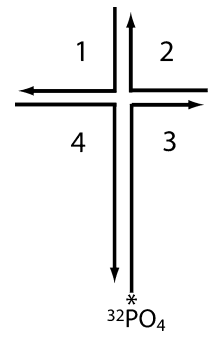


Fig. 1 Schematic formation of a cruciform structure. The *arrows* indicate the 3′-termini of the annealed oligonucleotides 1, 2, 3 and 4; the *asterisk* indicates the radioactive $^{32}PO_4$ group on oligonucleotide 3

3. T4 Polynucleotide Kinase (NEB).
4. 7.5 M ammonium acetate.
5. 95 and 70 % ethanol.
6. TMS annealing buffer: 100 mM NaCl, 10 mM Tris–HCl, pH 7.5, 10 mM $MgCl_2$.
7. 2× Binding buffer: 16 % Ficoll, 200 mM NaCl, 20 mM HEPES, pH 7.9, 10 mM KCl, 2 mM EDTA, 2 mM spermidine, 1 mM DTT.
8. TBS:10 mM Tris–HCl pH 7.5, 100 mM NaCl.
9. TBE (10×) prepare 1 L by mixing 108 g Tris-base, 55 g boric acid, 40 ml 0.5 M EDTA, pH 8.0.
10. 40 % Actylamide stock solution (38 parts acrylamide/2 parts bisacrylamide).
11. 0.1 % Xylene cyanol.
12. Gel loading buffer (10×) 0.25 % bromophenol blue, 0.25 % xylene cyanol, 25 % Ficoll 400.
13. Prepare multiple 0.5 ml Eppendorf tubes with pierced bottom (make several holes with an 18–21-gauge syringe needle) and fill the bottom of the tubes with glass wool (keep stock in 70 % ethanol). Autoclave and keep in a sterile jar until use.
14. For competition experiments, double-stranded linear DNA containing the target DNA sequence of interest, stock solution at least 200–500 μg/ml in ddH_2O.
15. For antibody supershift, specific antibodies against the protein of interest in serum dilutions ranging from undiluted to 1:1000 in TBS.

3 Methods

3.1 5′-End Labeling of Oligonucleotide 3

1. Take 2 μl (20 pmol) oligo 3 and add 2 μl 10× Polynucleotide Kinase buffer (NEB) 10 μl [γ-^{32}P]-ATP, 5 μl H_2O, and 1 μl T4 Polynucleotide Kinase (~10 units) (NEB).
2. Incubate for 1 h at 37 °C.
3. Add 10 ml 7.5 M ammonium acetate (final concentration 2.5 M).
4. Add 270 μl 95 % ethanol.
5. Leave for 5 min, at RT.
6. Spin in an Eppendorf centrifuge 5 min at 15,000 × *g*.
7. Carefully discard supernatant into radioactive waste, add 300 μl 70 % ethanol.
8. Spin for 1 min at 15,000 × *g*.
9. Carefully discard supernatant into radioactive waste.

3.2 Annealing of the Oligonucleotides and Isolation of the Labeled Cruciform

1. Add 2 µl (20 pmol) each of oligos 1, 2, and 4 to precipitated, labeled oligo 3.
2. Add 25 µl TMS and mix thoroughly to redissolve oligo 3.
3. Take a small aliquot to determine approximate specific activity, that is, total Cerenkov cpm/20 pmol oligo 3.
4. Place in an aluminum heating block at 90 °C, switch off heating immediately, and insulate by covering with a Styrofoam box. Leave for 3 h to anneal cruciform.
5. Alternative method: Leave for 3 min at 90 °C, then 10 min at 68 °C and 30 min at 37 °C. This can be easily performed in a thermal cycler.
6. Add 3 µl 10× gel loading buffer.
7. Load into a 1-cm-wide pocket on a 1-mm-thick, 20-cm-long 6.5 % polyacrylamide gel in 1× TBE.
8. Run at 10 V/cm for 2–3 h until bromophenol blue has migrated about 14 cm.
9. Remove the upper plate of the gel, cover with Saran wrap, and autoradiograph for 30 s to 1 min. Ideally, use fluorescent ink markers to allow realignment of film to gel. Alternatively, use one corner of a radiography cassette to align the film with the gel plate during exposure.
10. Realign the developed film under the gel on a transilluminator and excise the cruciform band using a scalpel. The cruciform migrates slightly below the xylene cyanol band on a 6.5 % gel, *see* Fig. 2 and **Note 1**.

3.3 Extraction of Cruciform

1. Place the cut gel fragment containing the cruciform in a 0.5 ml Eppendorf tube. Pierce several holes through the bottom of the tube with an 18–21-gauge syringe needle. Place the tube in a 1.5 ml Eppendorf tube without lid.
2. Centrifuge 20–30 s or until the gel passes through the holes into the bottom tube forming a gel pellet.
3. Add 300 µl TMS to the pellet to form a slurry.
4. Seal the tube and leave overnight at 4 °C.
5. Put a previously prepared glass wool-bottomed, pierced, autoclaved 0.5 ml tube (see Subheading 2, **item 13**) in an intact 1.5 ml Eppendorf without lid.
6. To recover eluate, transfer all of the gel slurry into the 0.5-ml tube using a 1-ml pipette with a large tip opening (cutoff), and centrifuge for 20 s.
7. Adjust the eluate volume to 400 µl with TMS.

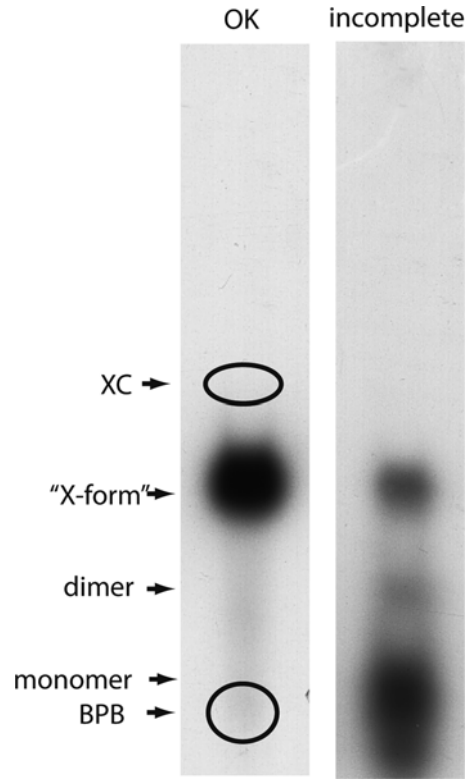


Fig. 2 Migration of the cruciform on the preparative gel. The cruciform structure was resolved on a 6.5 % polyacrylamide gel and exposed as described in Subheading 3. *Left panel*: position of the cruciform (X-form) relative to the dyes ("XC" xylene cyanol and "BPB" bromophenol blue), indicated by *black ellipses*. *Right panel*: an example of incomplete annealing. Positions of the monomer and dimer forms are indicated on the *left side* of the figure

The eluted cruciform DNA should be at ~50 fmol/μl and nearly all the radioactivity should have been eluted. Cerenkov count a 2 μl aliquot in order to calculate the concentration of cruciform using the specific activity calculated from Subheading 3.2 **step 3**. Stored at 4 °C, the labeled cruciform is stable for at least a week. You will need about 100 fmol (2 μl) per mobility shift assay.

3.4 Cruciform Mobility Shift Assay

Each mobility shift reaction is performed in 10 μl consisting of 5 μl of 2× binding buffer, 2 μl (100 fmol) of cruciform DNA in TMS, and 3 μl of TBS containing varying amounts of the proteins to be assayed.

1. Prepare a 6.5 % polyacrylamide gel (38 parts acrylamide/2 parts bisacrylamide) at least 15–20 cm long and 1 mm thick in 0.5× TBE, and use a 0.5 % TBE as running buffer. Pre-run the gel for 1 h at 11 V/cm.

2. Prepare a stock mix of 2× binding buffer (5 µl × number of reactions plus one) and cruciform (2 µl (100 fmol) × number of reactions plus one).
3. Prepare dilutions of the protein to be analyzed in Eppendorf tubes. Complete the volume of the protein in each tube to 3 µl with TBS. Place 3 µl TBS in a tube to serve as a negative control. Since the K_d of the expected complexes is typically less than µmolar, start with protein amounts ranging from 2 to 30 pmol.
4. Add 7 µl binding buffer/cruciform mix, **step 2** above, to each protein dilution.
5. Incubate 10–30 min at room temperature.
6. Add 0.5 µl 0.1 % xylene cyanol to each sample and load onto the gel.
7. Electrophorese for 3–4 h at 11 V/cm.
8. Transfer the gel onto a sheet of Whatman 3MM paper and cover with Saran wrap.
9. Dry the gel for 30 min at 85 °C.
10. Expose the gel to radiography film or use a commercial phosphorimaging device to detect cruciform and analyze the results. A typical example of the electrophoretic separation is shown in Fig. 3, *see* **Notes 2** and **3**.

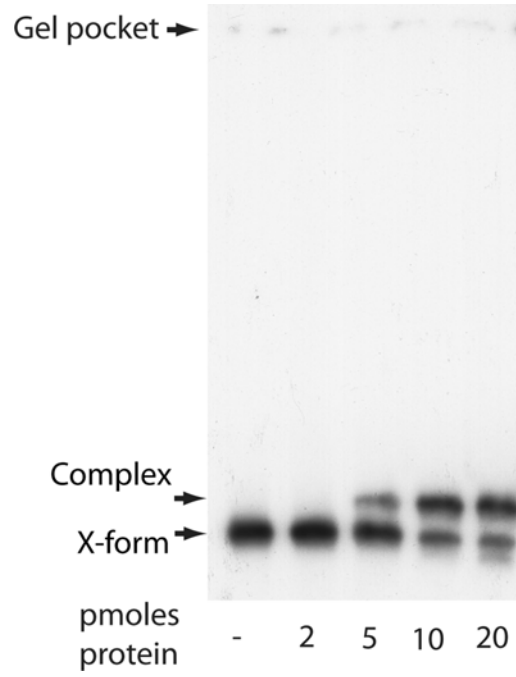


Fig. 3 A typical example of cruciform shift assay. About 100 fmol of cruciform was incubated with the indicated amounts of HMGB-box1 from UBF, *see* Subheading 3. The protein–DNA complex (Complex) and the naked cruciform (X-form) are indicated

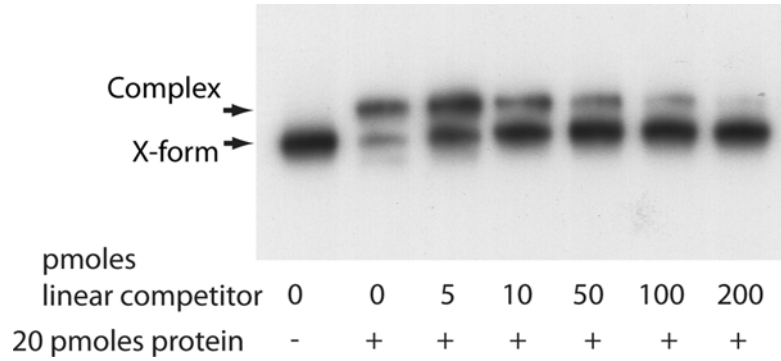


Fig. 4 Competition with linear target sequence. The cruciform structure was incubated with 20 pmol of UBF HMGB-box 1(as in Fig. 3) and with increasing amounts of linear human rDNA promoter UCE fragment, *see* [8]. In this case the cruciform shift is efficiently competed with 200 pmol of the linear promoter fragment. Complex and X-form as in Fig. 3

3.5 Competition with Linear DNA

The affinity of the protein for linear relative to cruciform DNA can be determined in a simple competition assay, *see* Fig. 4. Typically, a one- to tenfold molar excess of the linear DNA fragment over the protein is required. Add 1 μl of the double-stranded linear fragment to the cruciform DNA before mixing with the protein sample in **step 3** of Subheading 3.4, then proceed as described in that section. A protein concentration that yields less than complete shifting of the cruciform in the absence of linear DNA must be used if the relative affinity of the protein for linear DNA is to be determined.

3.6 Supershifting with a Specific Protein Antibody

In the case of impure protein samples, it may be necessary to determine the identity of the protein that is responsible for the cruciform shift. This can sometimes be achieved by "upshifting" the cruciform–protein complex by the addition of an antibody before electrophoretic analysis. Add 1 μl of an appropriate range of antibody dilutions to the samples just before applying them to the gel, then proceed as described in that section. An example of such an upshift assay is shown in Fig. 5.

4 Notes

1. Annealing of the cruciform is not complete (*see* Fig. 2b, right panel). Usually, the annealing is so efficient that only the completed cruciform, containing all four oligos, is visible on the gel. In some cases, however, smaller structures containing one or two oligos may be present, running closer to the bromophenol blue. If the cruciform is not the major product:
   (a) Repeat the annealing after verifying the identity and quality of oligonucleotides.

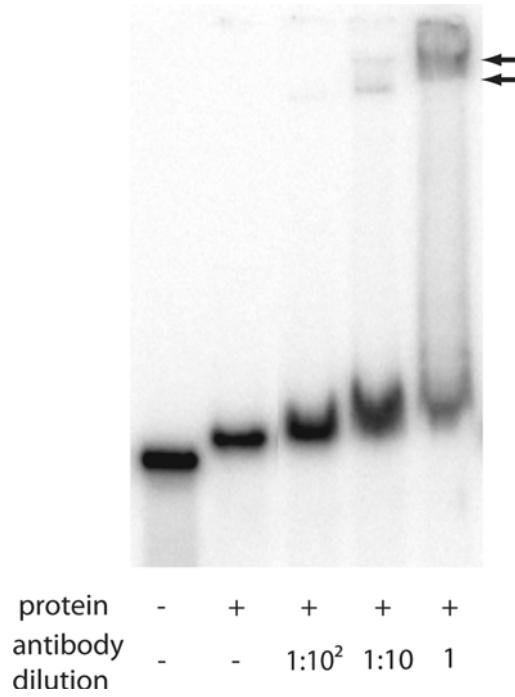


Fig. 5 Specific antibody "supershift." The cruciform structure (100 fmol) was incubated with 20 pmol of UBF HMGB-box1 and 1 μl diluted polyclonal anti-UBF antibody as indicated. An increasing large "supershift" is observed with an increase in antibody concentration, eventually producing antibody–protein–DNA complexes high in the gel indicated by *arrows*

(b) Check the concentration of your oligonucleotides. Equimolar amounts must be used.

(c) The gel may have been run too hot. Check the voltage; it should not exceed 11 V/cm.

2. More than one shifted band is visible on the gel. There may be a second cruciform-binding protein in the protein preparation. A cruciform in some cases may bind more than one protein molecule. *See* also **Note 3**.

3. Certain protein–DNA complexes migrate faster than the naked cruciform (downshift). Depending on buffer conditions, cruciforms may adopt alternative tertiary folds, affecting their mobility during electrophoresis. For example, HMGB-box 2 of UBF under standard buffer conditions gave a downshift as well as an upshift. However, addition of an extra 3 mM $MgCl_2$ during the binding reaction eliminated the downshift, *see* Fig. 6. It is possible that acrylamide concentration or acrylamide/bisacrylamide ratio in the gel may also affect relative migration of cruciform structural isomers.

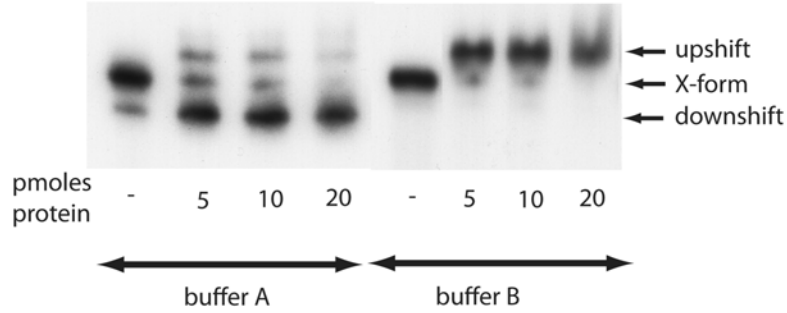


Fig. 6 Differential migration of the protein-cruciform DNA complex depending on buffer conditions. Increasing amounts of UBF HMGB-box 2 were incubated with the cruciform using two different binding buffers; buffer A, the standard binding reaction (2 mM $MgCl_2$ final concentration) and buffer B, as buffer A but giving final concentrations of 5 mM $MgCl_2$ and 0.3 mM ATP. Positions of the free cruciform (X-form), the "upshift" and the "downshift" are indicated with *arrows*

Acknowledgments

This work was supported by an operating grant from the Canadian Institutes of Health Research.

References

1. Bianchi ME, Beltrame M, Paonessa G (1989) Specific recognition of cruciform DNA by nuclear protein HMG1. Science 243:1056–1059
2. Ferrari S, Harley VR, Pontiggia A, Goodfellow PN, Lovell-Badge R, Bianchi ME (1992) SRY, like HMG1, recognizes sharp angles in DNA. EMBO J 11:4497–4506
3. Lilley DM (1992) DNA–protein interactions. HMG has DNA wrapped up. Nature 357:282–283
4. Van de Wetering M, Clevers H (1992) Sequence-specific interaction of the HMG box proteins TCF-1 and SRY occurs within the minor groove of a Watson-Crick double helix. EMBO J 11: 3039–3044
5. Kuhn A, Stefanovsky V, Grummt I (1993) The nucleolar transcription activator UBF relieves Ku antigen-mediated repression of mouse ribosomal gene transcription. Nucleic Acids Res 21: 2057–2063
6. Bazett-Jones DP, Leblanc B, Herfort M, Moss T (1994) Short-range DNA looping by the Xenopus HMG-box transcription factor, xUBF. Science 264:1134–1137
7. Stefanovsky VY, Bazett-Jones DP, Pelletier G, Moss T (1996) The DNA supercoiling architecture induced by the transcription factor xUBF requires three of its five HMG-boxes. Nucleic Acids Res 24:3208–3215
8. Stefanovsky VY, Pelletier G, Bazett-Jones DP, Moss T (2006) ERK modulates DNA bending and enhancesome structure by phosphorylating HMG1-boxes 1 and 2 of the RNA polymerase I transcription factor UBF. Biochemistry 45:3626–3634
9. Zannis-Hadjopoulos M, Yahyaoui W, Callejo M (2008) 14-3-3 cruciform-binding proteins as regulators of eukaryotic DNA replication. Trends Biochem Sci 33:44–50

Chapter 13

Individual and Sequential Chromatin Immunoprecipitation Protocols

Mayra Furlan-Magaril and Félix Recillas-Targa

Abstract

DNA regulatory elements nucleate the interaction of several transcription factors in conjunction with ubiquitous and/or tissue-specific cofactors in order to regulate gene expression making it relevant to determine the profiles of cohabitation of several proteins on the chromatin fiber. Chromatin immunoprecipitation (ChIP) has been broadly used to determine the profile of several histone posttranslational modifications as well as transcription factor occupancy in vivo. However, individual ChIP does not resolve whether the epitope under study is present at the same time on a given genomic location. Here we describe the ChIP-re-ChIP assay that represents a direct strategy to determine the in vivo co-localization of proteins or histone posttranslational modifications in a chromatinized template on the basis of double and independent rounds of immunoprecipitation with high-quality ChIP-grade antibodies.

Key words Chromatin, Protein-DNA interaction, Immunoprecipitation, Histone modifications, GATA-1, YY-1, Telomeric position effect

1 Introduction

The chromatin status and the in vivo association of transcription factors to DNA regulatory regions and RNA are essential to understand the transcriptional regulatory circuits that orchestrate cellular processes such as cell proliferation and differentiation among others. Chromatin immunoprecipitation (ChIP) has proven to be a powerful assay to characterize DNA-protein interactions as well as histone posttranslational modifications in vivo. The principle of this strategy relies on the use of antibodies against the factor of interest to immunoprecipitate a protein-DNA complex stabilized by a crosslinking agent, usually formaldehyde. In recent studies, alternative cross-linkers such as EGS or DGS have been used in combination with formaldehyde to recover larger protein-chromatin complexes. Formaldehyde is the best option for molecules that interact directly with DNA. EGS (ethylene glycol bis[succinimidylsuccinate]) and DGS (disuccinimidyl glutarate)

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_13,

cross-linkers form covalent bonds between residues of molecules as far as 16.1 and 7.7 Å, respectively, allowing the detection of large protein complexes interacting indirectly with DNA [1, 2].

Irrespective of the cross-linker chosen, ChIP does not provide the information regarding if two factors or modifications of interest occupy the chromatin fiber at a given time. A modification of the individual ChIP protocol named here ChIP-reChIP or sequential ChIP is the protocol of choice to determine whether two proteins or histone posttranslational modifications coexist in the chromatin template at the same time [3–5]. Sequential ChIP is based on the principle that chromatin and associated transcription factors and/or cofactors are first immunoprecipitated with a specific antibody and the eluted material is then subjected to another immunoprecipitation with a second antibody. Importantly, this second antibody recognizes a protein that is suspected to coexist near or forms a direct or indirect complex with the first immunoprecipitated antigen. Different sized complexes can be obtained depending on the cross-linker selected. Finally, the crosslinks are reverted and the recovered DNA is purified and constitutes the genomic template for either library preparation in case of massive parallel sequencing or PCR amplification using specific primers for the genomic region of interest.

Although ChIP-reChIP can suggest the cohabitation of two proteins in the genome and their probable physical interaction, direct protein-protein contacts cannot be addressed by sequential ChIP assay; for such an aim alternative strategies such as standard Co-IP or in vitro pull down can be used.

1.1 Chromatin Status

The applicability of sequential ChIP has been described in a variety of cellular contexts including embryonic stem (ES) cells [6, 7]. ES cells are pluripotent cells present in the early mouse embryo that undergo intense chromatin changes that involve histone modifications and chromatin movement inside the nucleus [8, 9]. It has been shown through sequential ChIP that on particular locations of the ES cell genome there are "bivalent chromatin domains." These refer to the coexistence of repressive marks (H3K27me3) and active marks (H3K4me3) on the same genomic region [10]. This coexistence is proposed to prepare the genome of ES cells in order to make rapid epigenetic decisions on the transcriptional status (activation or repression). Other studies are also examples of the applicability of ChIP-reChIP to detect diverse epitopes on chromatin including the co-localization of histone variants [11], protein posttranslational modifications as SUMOylation [12] and hormone receptor binding [13], among others.

The analysis of individual ChIP and sequential ChIP assays can be performed in a quantitative manner through the use of real-time PCR or duplex-PCR to analyze a particular set of genomic regions [10, 14, 15] or be prepared for massive parallel sequencing to obtain genome-wide data [16, 17].

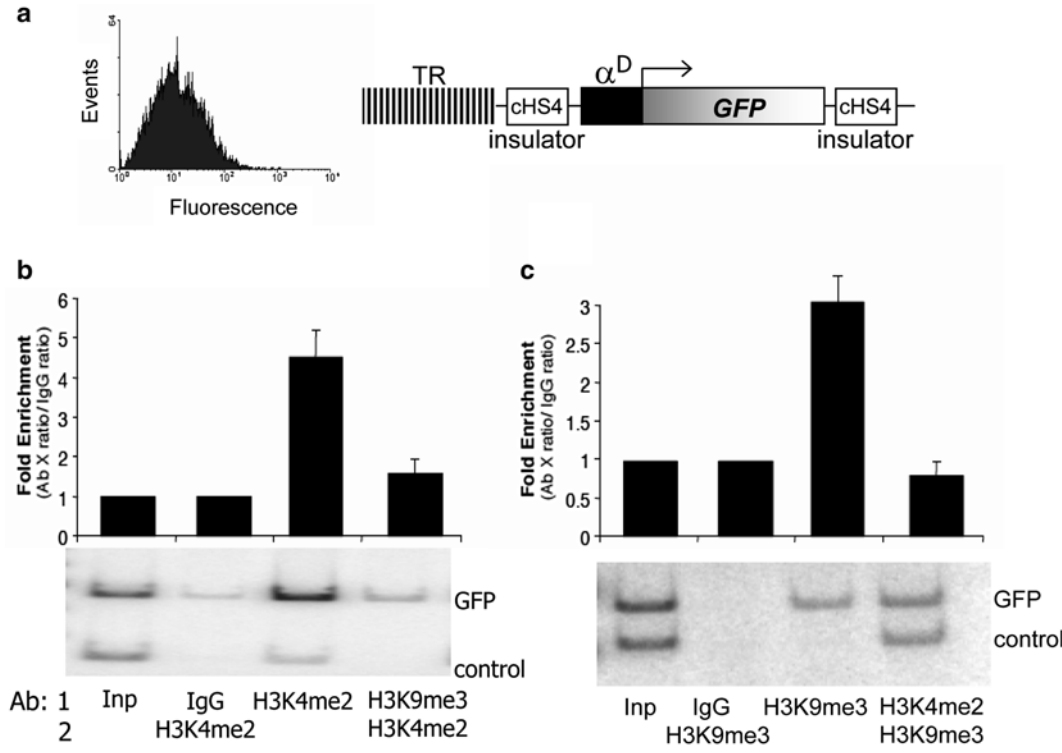


Fig. 1 Re-ChIP analysis by duplex PCR. In this experiment the coexistence of H3K4me2 and H3K9me3 in the body of a *GFP* reporter gene was analyzed. HD3 cells were stably transfected with the plasmid shown in the diagram. The plasmid contains *GFP* reporter gene under chicken α^D globin gene promoter; additionally, the transgene is flanked by the 5′HS4 β-globin insulator and contains a telomeric repeat fragment. This plasmid was used to evaluate the protection against telomeric position effect of the 5′HS4 insulator [15]. (**a**) Fluorescent cytometry analysis of the *GFP-expressing* clone used for re-ChIP experiments. (**b**) Re-ChIP using H3K4me2 for the first round of IP. The control region used corresponds to a heterochromatic region upstream of the β-globin gene domain in which this mark is absent. As can be seen in the graph and gel, there is no re-ChIP signal even if the H3K4me2 alone is present. (**c**) Re-ChIP using H3K9me3 for the first round of IP. The control region used corresponds to the chicken folate receptor hypersensitive site HSA known to be in an open chromatin conformation that lacks H3K9me3. There is no re-ChIP signal even the H3K9me3 is present. The previous experiments show that the two marks are not localized in the same region. In consequence there are two different cell populations, one with each histone modification

In our laboratory, we have systematically used radioactive duplex PCR to analyze the co-localization of antagonistic histone covalent modification over stably integrated transgenes in cultures cells, in the presence or absence of chromatin insulators ([15], Fig. 1). Duplex PCR allows the normalization of immunoprecipitation enrichments and the comparison, in the same genomic location, of the relative abundance of specific histone posttranslational modifications [18]. Particular attention is needed in the design of these control primers, since they should correspond to known genomic region where the chromatin state is contrary to the histone mark under study. For example, if we analyze H3 acetylation over

a region (corresponding to a permissive chromatin feature), a set of control primers are needed to amplify a known region with heterochromatic features (corresponding to a nonpermissive chromatin feature) and vice versa. Of course, quantitative real-time PCR is also a valuable alternative.

1.2 In Vivo Determination of Protein Complex Formation

Performing sequential ChIP, we have recently determined that GATA-1 and YY1 transcription factors co-localize in vivo at the chicken α-globin enhancer ([4], Fig. 1). As previously mentioned, sequential ChIP does not allow the determination of direct protein-protein interactions, therefore whole extract co-immunoprecipitations and pull-down experiments were further required to demonstrate a physical contact between them. These results do not discard the interaction of other factors, in particular, specific cofactors with the capacity to recruit chromatin-remodeling activities. This kind of association has been recently demonstrated for the nuclear protein NLI/Ldb1 (the human homolog of the *Drosophila melanogaster* protein ChIP), which was not known to possess enzymatic or DNA-binding activity [5, 19]. Sequential ChIP demonstrated that NLI/Ldb1 is able to form a complex with LMO2, SCL and GATA-1 that contributes to β-globin regulation, but most importantly to facilitate long-range interaction between the β-globin locus control region and the target globin gene by chromatin loop formation [5]. This sequential ChIP experiments were complemented with a ChIP assay followed by a chromosome conformation capture (3C) assay demonstrating that NLI is responsible for long distance interactions [5]. These are two examples supporting the experimental contribution of the ChIP-reChIP assay to the in vivo determination of protein location [3, 5, 14, 20].

Based on the vast amount of data showing that genomic regulatory elements are bound by multiple proteins and a variety of histone modifications that contribute to its functional output, the ChIP-reChIP assay represents a very useful experimental tool to decipher the combinatorial of regulatory protein and chromatin modifications across the genome.

2 Materials

2.1 Cell Culture

1. HD3 cells are chicken erythroblasts arrested in a CFU stage via infection with the avian erythroblastosis virus (AEV). HD3 cells are cultured in DMEM supplemented with 8 % fetal bovine serum, 2 % chicken serum, and 1× penicillin/streptomycin at 37 °C, 5 % CO_2, in a 150 mm dish.

2.2 ChIP-ReChIP

1. PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, and 2 mM $KHPO_4$ pH 7.2).
2. PBS/2 % FBS/1 mM PMSF.

3. Formaldehyde 11 % (7.45 mL 37 % formaldehyde, 0.1 M NaCl, 1 mM EDTA, 0.5 mM EGTA, and 50 mM HEPES). Alternatively DGS or EGS can be used as cross-linkers together with formaldehyde (Thermo Scientific).
4. 2.5 M Glycine (18.767 g in 100 mL sterile water).
5. Lysis buffer (1 % SDS, 5 mM EDTA, 50 mM Tris–HCl pH 8.1; supplement with protease and deacetylase inhibitors before using as follows: 1 mM PMSF, 100 μM leupeptin, 0.3 μM aprotinin, 1 μm pepstatin A, 10 μm bestatin, and 1 mm sodium butyrate).
6. Dilution buffer (1 % Triton X-100, 2 mM EDTA, 20 mM Tris–HCl pH 8.1, 150 mM NaCl; supplemented with protease inhibitors before using).
7. Wash buffer I (0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris–HCl pH 8.1; supplemented with protease inhibitors before using).
8. Wash buffer II (0.1 % SDS, 1 % Triton X-100, 2 mM EDTA, 500 mM NaCl, 20 mM Tris–HCl pH 8.1; supplemented with protease inhibitors before using).
9. Wash buffer III (0.25 M LiCl, 1 % NP40, 1 % deoxycholate, 1 mM EDTA, 10 mM Tris–HCl pH 8.1; supplemented with protease inhibitors before using).
10. TE (10 mM Tris–HCl pH 8, 2 mM EDTA; supplemented with protease inhibitors before using).
11. Washing beads buffer (0.1 mM $NaHCO_3$, 1 % SDS; add 10 μg/μL proteinase K and 10 μg/μL RNase A before use).
12. Protein A and G (Amersham): Once beads (protein A + G) are preabsorbed with BSA and salmon sperm DNA, store them at 4 °C for up to 6 months. Alternatively Sepharose A and G Dynabeads (Invitrogen) can be used. In this case, washes are simplified by using a magnet for bead collection instead of centrifugation throughout the protocol.
13. Bovine serum albumin fraction V (BSA).
14. Salmon sperm DNA.
15. "MiniElute purification Kit" (Qiagen).
16. Phenol:chloroform:isoamyl alcohol (25:24:1).
17. Antibodies against H3K4me2 and H3K9me3 (Dr. Thomas Jenuwein, Max Planck Institute of Immunology and Epigenetics, Freiburg, Germany), rabbit α-IgG (Zymed) (Fig. 1), antibodies against GATA-1 (H-200 Santa Cruz Biotechnology), and YY1 (H-414 Santa Cruz Biotechnology) (Fig. 2).
18. Dithiothreitol (DTT).

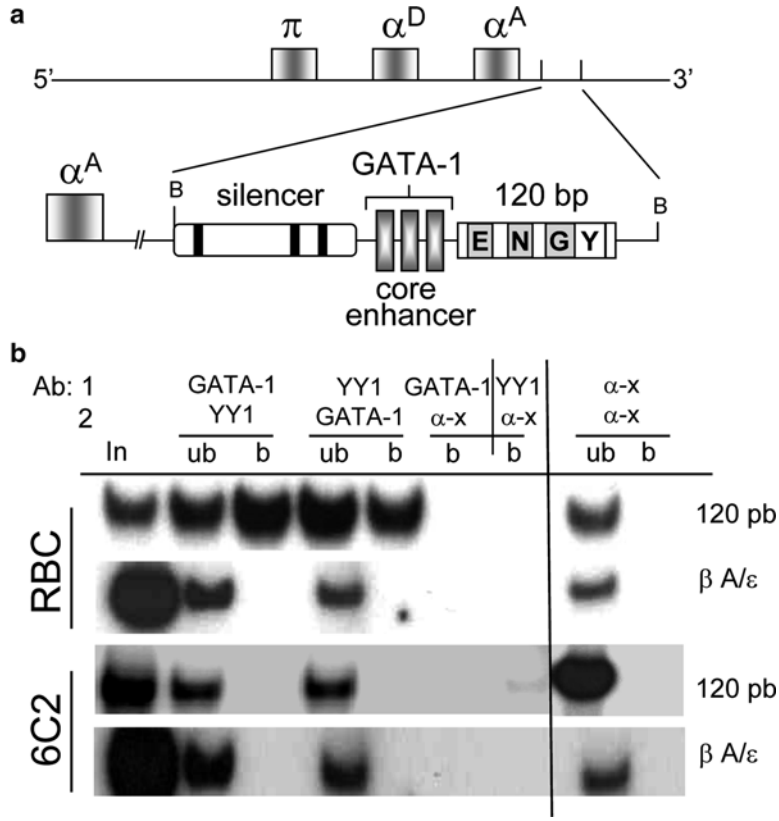


Fig. 2 Re-ChIP analysis by qualitative PCR. In this experiment we analyze the coexistence of GATA-1 and YY1 in a 120 bp region of the 3′enhancer of the chicken α-globin gene domain. These factors have been shown to modulate enhancer's activity in an erythroid stage-specific manner [4]. (**a**) Diagram of the α-globin gene domain and the 3′enhancer element. The core enhancer presents three GATA-1-binding sites. In addition, a 120 bp conserved sequence downstream of the core enhancer has an EKLF (*E*), NF-E2 (*N*), and a fourth GATA-1 (*G*) and YY1-binding site (*Y*). (**b**) Re-ChIP of GATA-1 and YY1. As can be seen, both factors are binding the 120 bp in the mature red blood cell (RBC) stage while in 6C2 cells, which represent a pre-erythroblastic stage in erythroid differentiation, they are not. The above results show that binding of both factors to the 120 bp is stage dependent and these correspond to their influence on enhancer action over the globin genes [4]. Control region corresponds to the $\beta^{A/}\varepsilon$ enhancer in the β-globin gene domain

2.3 Duplex PCR and Acrylamide Gel

1. Taq Polymerase.
2. dNTPs.
3. Specific oligonucleotides.
4. [α-^{32}P] dCTP.
5. 40 % Acrylamide.
6. TBE buffer.
7. APS and TEMED.

8. Phosphor-screen (Amersham).
9. Typhoon Scanner.
10. Image Quant program.

3 Methods

The purpose of ChIP-reChIP assay is to evaluate the coexistence of two different factors or histone modifications in a given genomic region. A ChIP assay alone is unable to distinguish between two factors coexisting in the same DNA region and having two different cell populations with each one. The ChIP-reChIP experiment basically consists in performing a ChIP assay with two rounds of immunoprecipitation before DNA recovery. For simplicity in the following protocol, we will refer to antibody A and B for each of the putative factors to analyze. Two examples are presented (Figs. 1 and 2) in which different factors are analyzed in distinct genomic contexts.

There are three important controls one must consider before performing a ChIP-reChIP experiment:

1. The antibodies used for sequential immunoprecipitation should be inverted expecting to obtain the same PCR amplification product. In some cases it is possible for the immunoprecipitation enrichment to fail depending on the order in which the antibodies were used. This could suggest that one of the two antibodies is unable to access its epitope due to steric restraints, probably due to the associated proteins in the complex, or to a different relative abundance of the components under study that could impede their efficient detection [4].
2. Negative genomic controls for PCR amplification. A set of primers should be designed within a genomic sequence in which the interaction of the proteins under study is absent, thus expecting a clear absence of the PCR amplification product. For example, to evaluate a given regulatory element, a set of primers coming from an unrelated region (like an exon) or located on a distinct chromosome with a presumably different chromatin composition could represent a proper control.
3. Determine the specificity of the ChIP-reChIP assay, by performing the sequential immunoprecipitation with the antibodies under study in combination with an irrelevant antibody (i.e., nonspecific IgG).

3.1 ChIP-ReChIP Assay

DAY 1

Before starting, add protease and deacetylase inhibitors (1 mM PMSF, 100 μM leupeptin, 0.3 μm aprotinin, 1 μm pepstatin A, 10 μm bestatin, and 1 mm sodium butyrate) to lysis and dilution buffers.

1. Grow cells to an 80–90 % confluence in a 150 mm dish.
2. Recover the cells in a 50 mL Falcon tube.
3. Count the cells to assure an adequate and comparable number per assay (a minimum of 10×10^6 cells are needed per antibody). Chromatin for ten different antibodies is obtained from a cell culture performed in two 150 mm semi-confluent HD3 dishes ($10–20 \times 10^6$ cells) (*see* **Note 1**).
4. Centrifuge for 5 min ($200 \times g$) at 4 °C.
5. Wash twice with 20 mL PBS/2 % FBS/1 mM PMSF. Gently resuspend the pellet with a 25 mL pipette. Centrifuge each time for 5 min ($200 \times g$) at 4 °C.
6. Resuspend the pellet with 30 mL PBS. Cross-link by incubating cells with 1 % formaldehyde 10 min at room temperature (add 3 mL of a formaldehyde 11 % stock solution and mix gently by flipping the tube three times). If a different cross-linker as DSG is being used in addition to formaldehyde, add DSG to the cells first to a final concentration of 2 mM in PBS and incubate on a rocker at RT for 45 min. Centrifuge the cells, wash them once with ice-cold PBS, and then proceed to formaldehyde fixation.
7. Stop cross-link by adding glycine to a 125 mM final concentration (1.5 mL of a 2.5 M glycine stock solution). Gently mix by flipping the tubes.
8. Centrifuge for 5 min ($200 \times g$) at 4 °C.
9. Wash cells twice with 35 mL of cold PBS. Centrifuge each time for 5 min ($200 \times g$) at 4 °C.
10. Resuspend cells in 1 mL lysis buffer and transfer to a 15 mL Falcon tube. Incubate for 10 min on ice and proceed to chromatin sonication.
11. Sonicate chromatin to obtain fragments of a desired size depending on the expected PCR product (*see* **Note 2**). To obtain HD3 chromatin fragments between 300 and 500 bp sonicate with 35 % amplitude giving 12 pulses of 30 s each (Cole and Palmer ultrasonic processor). Be sure to keep the tip of the sonicator at the bottom of the tube without touching the tube walls while giving the pulse. Put the tube on ice for 2 min between pulses.
12. Transfer chromatin into a 1.5 mL tube and centrifuge ($15{,}000 \times g$) for 10 min at 4 °C.
13. Transfer the supernatant into a 15 mL Falcon tube. Take a 50 μL aliquot as INPUT and store it at –20 °C.
14. Dilute chromatin 1:10 with dilution buffer (*see* **Note 3**).
15. Incubate chromatin with 50 μL of previously blocked beads (see next) for 2 h at 4 °C in constant rotation (*see* **Note 4**).

16. Centrifuge (1000 × *g*) for 5 min at 4 °C.
17. Divide chromatin into 1 mL aliquots, one aliquot per antibody. This chromatin can be stored at 4 °C for up to 4 months.
18. For the first round of IP take four aliquots: (1) antibody A, (2) control IgG, (3) antibody B, and (4) control IgG (*see* **Note 5**).
19. Add 4 μg of antibody to each sample and incubate overnight at 4 °C in constant rotation (*see* **Note 6**).
20. Simultaneously, prepare sepharose beads as follows: Take 0.2 g of protein A and 100 μL of protein G in a final volume of 2 mL of 9:1 dilution buffer:lysis buffer. Preabsorb beads with 100 μg/mL BSA and 500 μg/mL salmon sperm DNA and rotate overnight at 4 °C.

DAY 2

Before starting add protease inhibitor to the lysis and dilution buffers as before.

21. Remove the pre-absorption mix and wash the beads twice with 2 mL of dilution buffer. Finally resuspend the beads in 2 mL 9:1 dilution buffer:lysis buffer. Store absorbed beads at 4 °C for up to 6 months.
22. Add 50 μL of preabsorbed beads per sample.
23. Rotate samples for 2–4 h at 4 °C.
24. Add protease inhibitors to the washing buffers I, II, III, and TE.
25. Centrifuge beads (800 × *g*) for 1 min in a conventional table centrifuge (Eppendorf). Store 50 μL of supernatant as "unbound" at −20 °C and remove the rest with a pipette (*see* **Note 7**).
26. Wash sequentially with 1 mL of washing buffer I, II, and III, rotating for 10 min at 4 °C. Centrifuge (800 × *g*) for 1 min between washes. Remove supernatant carefully with a pipette.
27. Wash twice with 1 mL TE as before.
28. Resuspend beads in 75 μL TE/10 mM DTT (*see* **Note 8**).
29. Elute immunocomplexes by incubating for 30 min at 37 °C.
30. Centrifuge (800 × *g*) for 2 min and transfer the supernatant into a clean 1.5 mL tube.
31. Dilute sample 20 times (to a final volume of 1.5 mL) with dilution buffer.
32. For the second round of IP use: (1) antibody B, (2) antibody B, (3) antibody A, and (4) antibody A. Take two new chromatin aliquots and incubate with (5) antibody A and (6) antibody B (*see* **Note 9**). Add 4 μg of antibody to each sample. Incubate overnight at 4 °C in constant rotation.

DAY 3

33. Proceed in the same way as before (**steps 21–26**).
34. Take the INPUT and UNBOUND samples out of the freezer.
35. Add 150 μL of washing beads buffer to each sample plus proteinase K and RNase A (5 μg/μL). Incubate for at least 4 h or overnight at 65 °C to reverse the cross-link.

DAY 4

36. Extract DNA using "MiniElute purification Kit" (Qiagen) following the manufacturer's instructions (*see* **Note 10**).
37. Elute DNA in 50 μL of sterile water. Dilute INPUT and UNBOUND samples up to 300 μL. Proceed to PCR reactions.

The obtained data could be analyzed using different approaches. Here we describe two of them: a semiquantitative one in which results are normalized using a control region in a duplex PCR reaction (Fig. 1) and a qualitative one in which no duplex PCR is performed and amplification signal of the immunoprecipitated (bound) chromatin is compared with the signal obtained using unbound sample as template (Fig. 2).

3.2 Semiquantitative Duplex PCR

In order to quantify the enrichment of a factor or histone mark (in this case antibodies A and B) in a given genomic region, a duplex PCR is performed. One PCR product corresponds to the region of interest and the other PCR product corresponds to a control region in which the evaluated mark is not expected, in order to normalize the data (Fig. 1).

1. Design the oligonucleotides to amplify the experimental and control regions (*see* **Note 11**).
2. Standardize radioactive duplex PCR with INPUT DNA. Be sure that the experimental/control ratio is near 1e and that amplification is at the linear range (*see* **Note 12**). Perform PCR reactions as indicated in Tables 1 and 2 (these conditions were the ones used for the experiment shown in Fig. 1).
3. Once the radioactive duplex PCR conditions are established proceed to perform reactions using the ChIP DNA as template. In this case (1) input, (2) IgG + A, (3) A, (4) B + A, (5) input, (6) IgG + B, (7) B, and (8) A + B.
4. Run samples in a 6 % acrylamide gel in 0.5 % TBE buffer at 150 V for 1 h and 45 min (*see* **Note 13**).
5. Dry the gel on 3 M paper for 2 h.
6. Expose gel on the phosphor-screen or film for 2 h.

Table 1
Radioactive PCR

Reactive	Amount
DNA (input or ChIP sample)	3 μL (of the 50 μL)
dNTPs (10 mM)	0.1 μL
dCTP [α ^{32}P] 10 mCi/mL	0.05 μL
Experimental oligos F + R (pool of 5 pmol each)	0.5 μL
Control oligos F + R (pool of 5 pmol each) (*see* **Note 15**)	0.5 μL
Buffer 10×	2 μL
Taq Polymerase	0.05 μL
H_2O	Up to 20 μL

Table 2
PCR cycle conditions

Cycle step	Conditions
Denaturation	95 °C for 5 min
Denaturation	95 °C for 50 s
Annealing	Tm for 50 s
Extension	72 °C for 50 s
Extension	72 °C for 5 min
Hold	25 °C

7. Reveal image on Typhoon Scan and analyze the results in Image Quant program using the volume report tool (*see* **Note 14**).
8. Once you have all values, normalize data by applying the following equation:

$$\text{Enrichment} = \frac{\left(\text{Experimental} / \text{Control}\right)\text{Ab}}{\left(\text{Experimental} / \text{Control}\right)\text{IgG}}$$

For each antibody you should use its corresponding IgG. For example, to normalize A + B antibodies sample use IgG + B. To normalize B + A antibodies sample use IgG + A and to normalize the sample A and sample B use input value.

9. Graph data (Fig. 1).

3.3 Qualitative PCR

In this case PCR reactions are performed independently in the experimental and control genomic regions and the signal obtained using the immunoprecipitated DNA as a template is compared with the signal obtained with the unbound DNA as template (Fig. 2). Specifications for PCR reactions and visualization are the same as described above.

4 Notes

1. The number of cells needed varies according to each cell type. Adjust cell number in order to obtain a good amount of chromatin (for at least ten independent antibodies) per experiment by measuring protein concentration (*see* **Note 3**). Keep cell number constant in each experiment as well as incubation and sonication conditions. Chromatin can be stored at 4 °C for up to 4 months.
2. In order to standardize sonication conditions make a pulse curve by taking a sample every 2 pulses (1–15 pulses). Treat chromatin with Proteinase K and RNase (500 μg/mL) for a minimum of 3 h at 65 °C. Purify DNA by phenol-chloroform extraction followed by precipitation. Load DNA in a 1 % agarose gel and visualize by ethidium bromide staining. Choose conditions in which the right size of chromatin fragments are generated.
3. Diluting 1:10 with dilution buffer makes a 10 mL solution in which 1 mL is used per antibody starting from 10 to 20×10^7 cells. If the cell type is changed, quantify protein concentration and dilute with dilution buffer + protease inhibitors to have a 1 μg/μL solution. Store it at 4 °C for up to 6 months. Take 400 μL (400 μg) per antibody.
4. Chromatin pre-clearance with absorbed beads reduces unspecific background. This step is not essential. If beads and IgG amplification signals are low, it can be skipped.
5. The IgG must be from the same species in which the antibody was generated.
6. 4 μg of antibody is an optimum amount to assure efficient immunoprecipitation. In order to reduce antibody, use 3 and 2 μg and assay for signal. Depending on the antibody 2 μg could be sufficient.
7. The aliquot of unbound material can be used in order to compare the bound one (chromatin that has specifically bound the antibody and is now precipitated with beads) (Fig. 2). Beads that now carry immunocomplexes are deposited at the bottom of the tube. When removing supernatant, be careful to avoid detaching the beads. When the washing buffers are added, beads will dislodge on their own. Do not vortex.

8. 10 mM DTT is used to elute immunocomplexes from beads. This can be done in 75–200 μL TE; consider this volume to dilute 20 times after incubation.
9. Re-ChIP assay enables to analyze if two given factors localize at the same genomic region. It is important to include independent immunoprecipitation for each one as controls. In the second round of IP include antibodies A and B in two separate chromatin samples.
10. DNA extraction could be done by phenol-chloroform and be followed by precipitation.
11. When designing experimental and control oligonucleotides consider that the control PCR product must come from a genomic region in which the analyzed factor is not expected. Also be sure both have similar Tm and that the size of the amplification products is adequate to separate them in an acrylamide gel.
12. In order to compare the amplification signal form the genomic region of interest against the control genomic region, the signal obtained with input DNA must be similar for both PCR products. Consider conditions in which experimental/control ratio of 0.8–1.2 is obtained. To have a real value of amplification signal, be sure that the PCR cycles used fall into the linear range. To do so, make an amplification curve using increasing amounts of input template (50, 100, 300, 500 ng). Amplification product has to increase linearly. If not, reduce PCR cycles until linear range is achieved.
13. Load samples in the following order: (1) input, (2) A, (3) IgG+A, (4) B+A, (5) input, (6) B, (7) IgG+B, and (8) A+B.
14. ImageQuant (GE Healthcare) is one of the programs available to analyze the obtained data. Take the square tool and make a little square around bands. Keep square area constant. Finally display the volume report and save data as an Excel file. Proceed to do the calculation. Other quantification tools can be used.
15. When performing qualitative analysis, control and experimental oligonucleotides are used in separate PCR reactions.

Acknowledgments

We would like to thank Georgina Guerrero and Fernanda Suaste Olmos for her excellent technical assistance. This work was supported by grants from the Dirección General de Asuntos del Personal Académico-UNAM (IN203811 and IN128464), Consejo Nacional de Ciencia y Tecnología, CONACyT (128464 and 220503).

References

1. Tavares L, Dimitrova E, Oxley D et al (2012) RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell 148:664–678
2. Mokry M, Hatzis P, Schuijers J et al (2012) Integrated genome-wide analysis of transcription factor occupancy, RNA polymerase II binding and steady-state RNA levels identify differentially regulated functional gene classes. Nucleic Acids Res 40:148–158
3. Hatzis P, Talinidis I (2002) Dynamics of enhancer-promoter communication during differentiation-induced gene activation. Mol Cell 10:1467–1477
4. Rincón-Arano H, Valadez-Graham V, Guerrero G et al (2005) YY1 and GATA-1 interaction modulate the chicken 3′-side α-globin enhancer activity. J Mol Biol 349:961–975
5. Song S-H, Hou C, Dean A (2007) A positive role for NLI/Ldb1 in long-range β-globin locus control region function. Mol Cell 28: 810–822
6. Surani MA, Hayashi K, Hajkova P (2007) Genetic and epigenetic regulators of pluripotency. Cell 128:747–762
7. Spivakov M, Fisher AG (2007) Epigenetic signatures of stem-cell identity. Nat Rev Genet 8:263–271
8. Misteli T (2007) Beyond the sequence: cellular organization of genome function. Cell 128: 787–800
9. Schneider R, Grosschedl R (2007) Dynamics and interplay of nuclear architecture, genome organization and gene expression. Genes Dev 21:3027–3043
10. Bernstein BE, Mikkersen TS, Xie X et al (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125:315–326
11. Nekrasov M, Soboleva TA, Jack C, Tremethick DJ (2013) Histone variant selectivity at the transcription start site: H2A.Z or H2A.Lap1. Nucleus 4:431–438
12. Balasubramaniyan N, Luo Y, Sun AQ, Suchy FJ (2013) SUMOylation of the farnesoid X receptor (FXR) regulates the expression of FXR target genes. J Biol Chem 288:13850–13862
13. Shankaranarayanan P, Mendoza-Parra MA, Walia M et al (2011) Single-tube linear DNA amplification (LinDA) for robust ChIP-seq. Nat Methods 8:565–567
14. Mikkersen TS, Ku M, Jaffe DB et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448:553–560
15. Rincón-Arano H, Furlan-Magaril M, Recillas-Targa F (2007) Protection against telomeric-position effects by the chicken cHS4 β-globin insulator. Proc Natl Acad Sci U S A 104: 14044–14049
16. Mendoza-Parra MA, Pattabhiraman S, Gronemeyer H (2012) Sequential chromatin immunoprecipitation protocol for global analysis through massive parallel sequencing (reChIP-seq). Protoc Exch. doi:10.1038/protex.2011.257
17. Martin D, Pantoja C, Fernández Miñán A et al (2011) Genome-wide CTCF distribution in vertebrates defines equivalent sites that aid the identification of disease-associated genes. Nat Struct Mol Biol 18:708–714
18. Escamilla-Del-Arenal M, Recillas-Targa F (2008) GATA-1 modulates the chromatin structure and activity of the chicken α-globin enhancer. Mol Cell Biol 28:575–586
19. Dorsett D (1999) Distant liaisons: long range enhancer-promoter interaction in *Drosophila*. Curr Opin Genet Dev 9:505–514
20. Lee TI, Johnstone SE, Young RA (2006) Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat Protoc 1:729–748

Chapter 14

Chromatin Endogenous Cleavage (ChEC) as a Method to Quantify Protein Interaction with Genomic DNA in *Saccharomyces cerevisiae*

Virginia Babl, Ulrike Stöckl, Herbert Tschochner, Philipp Milkereit, and Joachim Griesenbeck

Abstract

Chromatin endogenous cleavage (ChEC) is a technique which allows to monitor protein-DNA interaction in the nucleus of eukaryotic cells. In addition to mapping of genomic interaction sites ChEC may also yield quantitative information about the occupancy of proteins at their genomic target regions. Here, we provide a protocol for ChEC experiments in *S. cerevisiae*, downstream DNA analysis and quantification of ChEC-mediated degradation. The potential of the method is exemplified in ChEC experiments with RNA polymerase I and the yeast homolog of linker histone H1.

Key words ChEC, ChIP, Chromatin, Protein-DNA interaction, RNA polymerase I, Histone H1, Exponential phase, Stationary phase

1 Introduction

Methods for analysis of protein-DNA interactions in the living cell are important to understand eukaryotic nuclear processes. Chromatin immunoprecipitation (ChIP) has become an extremely valuable technique to investigate the association of a factor of interest with chromosomal loci [1]. ChIP is based on formaldehyde crosslinking of cells to stabilize transient interactions and co-purification of genomic DNA fragments with an immunoprecipitated protein of interest. ChIP relies on the availability of specific antibodies recognizing the formaldehyde treated protein which might be imbedded in a cross-linked nucleoprotein complex. Furthermore, the antibody-protein complex has to withstand the stringent washing conditions of the ChIP protocol. Usually, only a low percentage of DNA fragments is eventually recovered in the affinity-purified material. This enables straightforward relative quantification of protein-DNA interactions but might in most cases not allow to

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_14,

determine the absolute occupancy of a factor at a given genomic location.

Another method to analyze protein-DNA interactions is called chromatin endogenous cleavage (ChEC) [2]. ChEC relies on the expression of the protein of interest as a fusion protein with a C-terminal micrococcal nuclease (MNase). MNase is an endonuclease whose catalytic activity strongly depends on the presence of calcium ions. MNase fusion proteins are inactive when expressed in *S. cerevisiae* (hereafter called yeast), because the intracellular calcium levels are too low to activate the enzyme. In the presence of calcium, however, the MNase is activated and cleaves the DNA in its proximity. Cleavage events can be mapped to single genomic loci using Southern blotting and indirect end-labeling, or genome-wide after ligation-mediated PCR and microarray hybridization [3]. ChEC can be carried out upon calcium addition in vivo to permeabilized yeast cells or in vitro when calcium is added to nuclei isolated from formaldehyde-cross-linked cells [2]. DNA cleavage of the MNase fusion proteins is not necessarily due to direct binding to a target site, and may reveal information about the spatial localization of the fusion proteins and the cleaved DNA in the nuclear volume. Along these lines, MNase fused to ribosomal proteins has been used to probe the local tertiary structure within the ribosome [4].

Besides providing information about the genomic location of MNase fusion proteins, ChEC may also yield information about the occupancy of these proteins at the respective DNA fragments [5, 6]. The latter is achieved by monitoring the degree of degradation of a genomic DNA fragment under investigation by the MNase fusion protein in time course experiments. In contrast to most ChIP data, the percentage of DNA cleaved by MNase fusion proteins in ChEC experiments is significant and reveals that many DNA-binding factors associate with a large fraction of their respective target regions under certain conditions ([5, 6], *see* also Figs. 1 and 2).

Two major caveats are associated with the degradation analysis: (1) in vitro ChEC is carried out with nuclei isolated from formaldehyde treated cells and chemical crosslinking might in part impair the catalytic activity of the MNase. Thus, the degree of specific DNA degradation by the MNase fusion protein might provide only a lower limit of its actual occupancy at the DNA region under investigation. (2) MNase fusion proteins cross-linked to the genomic DNA might be released from the DNA in the course of ChEC and act as free MNases degrading genomic DNA nonspecifically. This has been observed for very abundant fusion proteins like histone-MNases [6]. Both of these issues are addressed in Subheading 3 provided below, describing the current ChEC protocol used in our group as well as downstream DNA analysis and quantification of ChEC-mediated degradation.

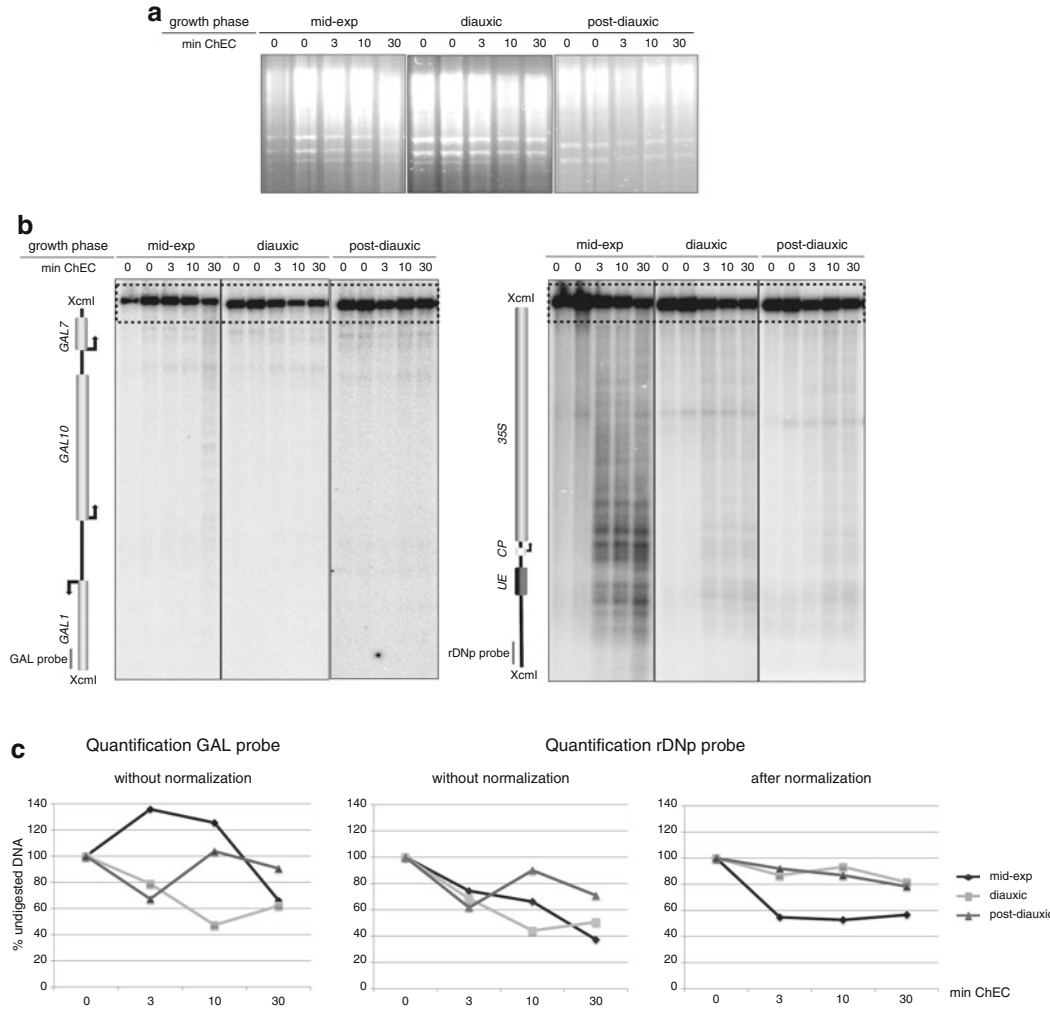


Fig. 1 Rpa190-MNase-mediated cleavage at two different genomic loci upon growth to stationary phase. Yeast strain y1717 [12], expressing Rpa190 as a fusion protein with C-terminal MNase from the endogenous chromosomal location, was grown in YPD + adenine. Cultivation was carried out for 48 h and samples were taken at different growth phases: at OD 0.5 (mid-exp), 9 h after OD 0.5 (diauxic), and 48 h after OD 0.5 (post-diauxic). Samples were subjected to ChEC for the times indicated above the individual lanes. Two control samples which were incubated in the absence of calcium together with the ChEC samples were included in the analysis (time of ChEC 0 min). DNA was isolated, digested with XcmI, separated in a 1 % agarose gel containing SYBR Safe, and visualized by fluorescence (**a**). (**b**) After transfer of the DNA from the gel to a membrane sequential Southern blot analysis using the indirect end-labeling technique [14] is performed with probes "GAL" hybridizing to the *GAL* locus (*left panel*) and "rDNp" hybridizing to the ribosomal DNA (rDNA) promoter region (*right panel*). Cartoons on the *left* of each autoradiograph show the genomic region investigated (*rectangles* depict the coding sequences of genes within the respective region, *arrows* mark transcription start sites; CP = core promoter; UE = upstream element; *small grey line* at the *bottom* of each cartoon indicates the region recognized by the radioactively labeled probe), *black dotted lines* frame the full-length DNA fragments used for quantification in (**c**). (**c**) Quantification of Rpa190-MNase mediated degradation. The signal intensities of the full-length fragment during the different ChEC time points (3, 10, and 30 min) were normalized to the mean of the signal intensities for the control samples, which was arbitrarily set to 100 % (graphs "without normalization"). Signals derived from the rDNA promoter region (*right panel*) were in addition normalized to the respective *GAL* probe signals to correct for DNA loading (graph "after normalization")

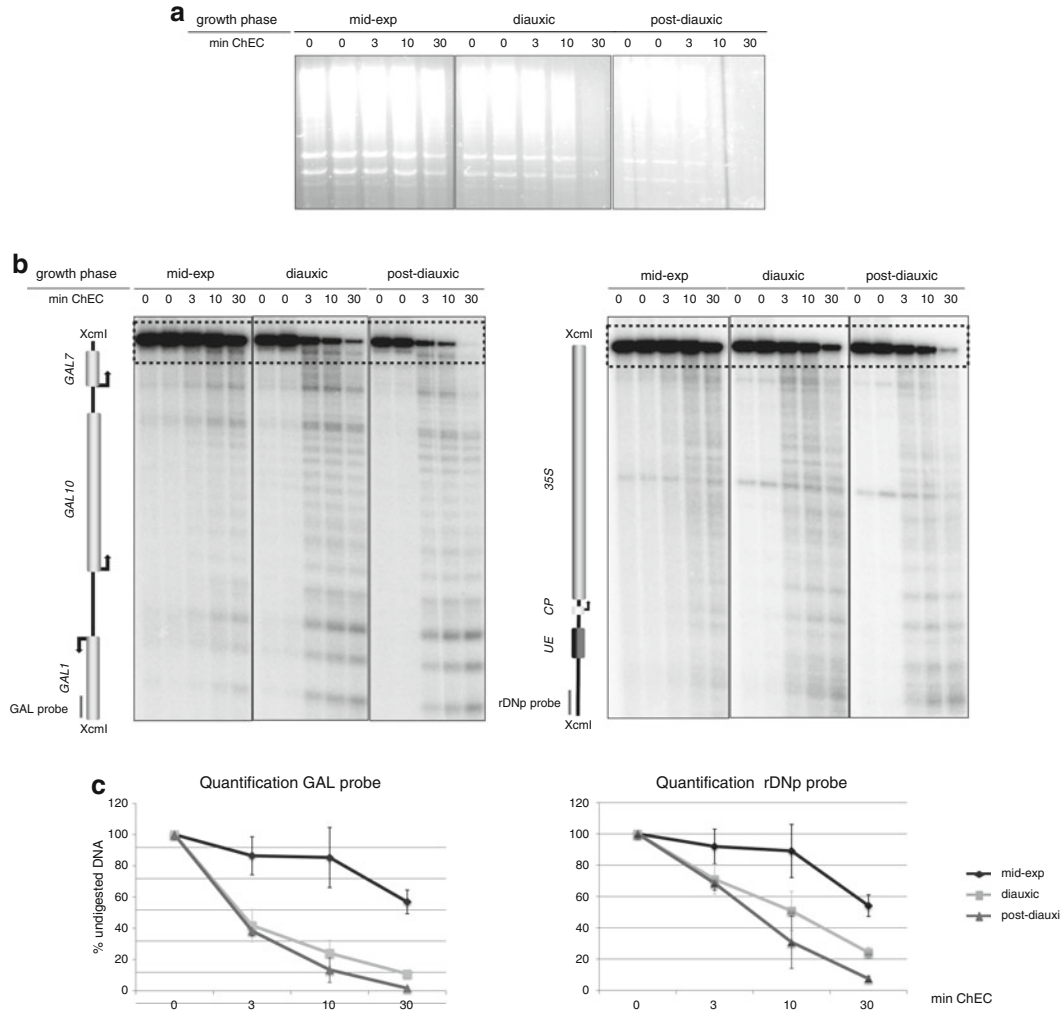


Fig. 2 Hho1-MNase-mediated cleavage at two different genomic loci upon growth to stationary phase. Yeast strain y1145 [15], expressing Hho1 as a fusion protein with C-terminal MNase from the endogenous chromosomal location, was grown in YPD + adenine. Cultivation and ChEC were carried out as described in Fig. 1. (**a**) Fluorescence image of the agarose gel. (**b**) Southern blot analysis as described in Fig. 1. (**c**) Quantification of Hho1-MNase-mediated degradation; the graphs show the average of two different experiments with error bars reflecting the standard deviation

To illustrate the potential of this method we show results of ChEC experiments with two different MNase fusion proteins, Rpa190, the largest subunit of RNA polymerase I which is moderately expressed (Fig. 1), and Hho1, the more abundant yeast homolog of linker histone H1 (Fig. 2). The only known genomic target for Rpa190 is the multi-copy gene for the large 35S ribosomal RNA (rRNA) precursor [7], whereas Hho1 binds genome-wide [8] (Figs. 1 and 2, compare results obtained with the *GAL* and rRNA gene locus-specific probes). Association of these two proteins with

the genomic DNA changes reciprocally when yeast cells grow into stationary phase. Rpa190 transcribes the 35S rRNA gene in exponentially growing cells but dissociates from its genomic target upon growth to stationary phase [9]. In contrast, Hho1 association with chromatin is weak in exponentially growing cells but strongly increases in stationary-phase cells [8]. This reported binding behavior is well reflected in the cleavage mediated by the respective MNase fusion proteins (compare Fig. 1 with Fig. 2). Whereas Rpa190-MNase-mediated cleavage at the 35S rRNA gene is strongly impaired when cells exit exponential growth, Hho1-MNase-mediated cleavage is strongly increased (compare Fig. 1c with Fig. 2c). It should be noted, that around 50 % of the 35S rRNA gene containing DNA fragments are digested in ChEC experiments with exponentially growing cells expressing Rpa190-MNase (Fig. 1c, graph on the right). This fits nicely with the observation that only about half of the 35S rRNA genes are transcribed by RNA polymerase I during exponential growth [6, 10, 11]. Interestingly, ChIP analyses of Rpa190-MNase from growing cells using the same strain analyzed in the above ChEC experiments leads to coprecipitation of only 2 % of the input DNA after immunoprecipitation [12].

2 Materials

All buffers, solutions, and media are prepared in water with a resistivity of 18.2 MΩ cm and a total organic content of less than five parts per billion. This is called "water" in the whole chapter.

2.1 Yeast Culture, Formaldehyde Fixation of Cells, and Isolation of Nuclei

1. Yeast peptone dextrose (YPD) (*see* **Note 1**).
2. Adenine hemisulfate salt (*see* **Note 2**).
3. Formaldehyde 37 % (stabilized with about 10 % methanol).
4. 2.5 M glycine, stored at room temperature (*see* **Note 3**).
5. IRN buffer: 50 mM Tris–HCl pH 8.0, 20 mM EDTA pH 8.0, 0.5 M NaCl, stored at 4 °C.
6. Spermine ≥99.0 % (GC), dissolved in water (1 M stock solution), stored in aliquots at −20 °C.
7. Spermidine tetrahydrochloride ≥98 % (TLC) dissolved in water (0.5 M stock solution), stored in aliquots at −20 °C.
8. Protease inhibitors (100×): Benzamidine (33 mg/mL), PMSF (17 mg/mL), dissolved in ethanol p.a., stored in aliquots at −20 °C.
9. Buffer A: 15 mM Tris–HCl pH 7.4, 80 mM KCl, 2–20 mM EDTA pH 8.0 (*see* **Note 4**), 2 mM EGTA pH 8.0, 0.5 mM spermidine, 0.2 mM spermine. Buffer should be stored at 4 °C. Protease inhibitors should be added freshly prior to use.

10. Glass beads (diameter 0.75–1.0 mm), stored at 4 °C.
11. Vibrax-VXR (IKA), rotary shaker.

2.2 Chromatin Endogenous Cleavage, DNA Extraction, and Restriction Enzyme Digest

1. Buffer Ag: 15 mM Tris–HCl pH 7.4, 80 mM KCl, 0.1 mM EGTA, stored at room temperature. After spermine and spermidine have been added to a final concentration of 0.5 mM and 0.2 mM, respectively, buffer should be stored at 4 °C. Protease inhibitors should be added freshly prior to use.
2. Thermomixer® Dry Block Heating Shaker (Eppendorf).
3. 0.1 M $CaCl_2$, stored at room temperature.
4. IRN buffer: 50 mM Tris–HCl pH 8.0, 20 mM EDTA pH 8.0, 0.5 M NaCl, stored at room temperature.
5. RNase A (20 mg/mL), dissolved in 50 mM Tris–HCl pH 8.0, 10 mM EDTA, stored at room temperature.
6. 10 % Sodium dodecyl sulfate (SDS), stored at room temperature.
7. Proteinase K (20 mg/mL), dissolved in 50 mM Tris–HCl pH 8.0 and 1 mM $CaCl_2$ and stored in aliquots at –20 °C.
8. Phenol/chloroform/isoamyl alcohol, stored at 4 °C in the dark.
9. Ethanol p.a., stored at –20 °C.
10. TE buffer: 10 mM Tris–HCl pH 8.0, 1 mM EDTA.
11. RNase A-TE buffer (0.05 mg/mL RNase A dissolved in TE-buffer), prepare freshly.
12. Restriction enzymes and buffers (New England Biolabs).

2.3 Agarose Gel Electrophoresis and Southern Blot

1. Tris-borate-EDTA (TBE): 89 mM Tris, 89 mM boric acid, 2 mM EDTA.
2. UltraPure™ Agarose (Invitrogen).
3. 10× DNA loading buffer: 0.25 % bromophenol blue (w/v), 0.25 % xylene cyanol (w/v), 40 % glycerol (w/v); stored at room temperature.
4. SYBR® Safe (Invitrogen) or ethidium bromide DNA stain.
5. FLA-3000 Imaging System (Fuji).
6. Denaturing solution: 0.5 M NaOH, 1.5 M NaCl, stored at room temperature.
7. Blotting solution: 1 M ammonium acetate, stored at room temperature.
8. Blotting paper MN 827 B (Macherey-Nagel).
9. Membrane Positive™ (3 × 0.3 m) (MP Biomedicals).
10. Parafilm stripes.
11. Paper towels.
12. TFL-35M Fluo-Link DNA Fixation and Visualization Transilluminator (Vilber Lourmat).

2.4 Probe Preparation, Membrane Hybridization, Washing, and Membrane Stripping

1. RadPrime DNA Labeling System (Invitrogen).
2. ProbeQuant™ G-50 Micro Columns (GE Healthcare).
3. UltraPure™ Salmon Sperm DNA Solution (10 mg/mL) (Invitrogen).
4. [α-^{32}P]dATP (specific activity: 111 TBq (3000 Ci)/mmol, concentration: 370 MBq (10 mCi)/mL).
5. Hybridization buffer: 0.5 M sodium phosphate pH 7.2, 7 % SDS, stored at room temperature, preheat to 65 °C prior to use.
6. 20× Saline-sodium citrate (SSC): 17.5 % sodium chloride, 8.8 % sodium citrate dihydrate, stored at room temperature.
7. Rinse buffer: 3× SSC, 0.1 % SDS, stored at room temperature, preheat to 65 °C prior to use.
8. Wash buffer 1: 0.3× SSC, 0.1 % SDS, stored at room temperature, preheat to 65 °C prior to use.
9. Wash buffer 2: 0.1× SSC, 0.1 % SDS, stored at room temperature, preheat to 65 °C prior to use.
10. Wash buffer 3: 0.1× SSC, 1.5 % SDS, stored at room temperature, preheat to 65 °C prior to use.
11. 20× Saline-sodium phosphate-EDTA (SSPE): 3.0 M sodium chloride, 0.2 M sodium phosphate pH 7.4, 0.02 M EDTA.
12. Stripping solution: 0.1× SSPE, 0.5 % SDS, stored at room temperature, preheat to 80 °C prior to use.

2.5 Quantification of Degradation

1. Multi Gauge v.3.0 (Fujifilm).
2. Microsoft Excel.

3 Methods

3.1 Yeast Culture, Formaldehyde Fixation, and Isolation of Nuclei

1. Yeast strains expressing proteins of interest fused to MNase are cultured at 30 °C to an OD_{600} of approximately 0.5 (mid-exponential phase). For the analysis of chromatin composition with transition to stationary phase the culture is incubated up to 144 h after reaching mid-exponential phase (*see* **Note 5**). Samples of roughly $0.5–1 \times 10^8$ cells (50 mL of culture grown to mid-exponential phase) are taken at different time points (*see* **Note 6**).
2. Formaldehyde (37 %) is added to the samples to a final concentration of 1 % (*see* **Note 7**) and the cell suspension is incubated at 30 °C for 15 min (*see* **Note 8**).
3. Excess formaldehyde is quenched by the addition of 2.5 M glycine to a final concentration of 125 mM, and the cell suspension is incubated at room temperature for at least 5 min.
4. Cells are pelleted in 50 mL reaction tubes for 5 min with $4200 \times g$ at 4 °C in a benchtop centrifuge and the supernatant

is discarded. Cells are suspended in 1 mL cold IRN buffer, transferred to a 1.5 mL microtube, and harvested for 2 min with 16,000 × *g* at 4 °C in a microcentrifuge, and the supernatant is discarded. Cells are frozen in liquid nitrogen and stored at −80 °C. Alternatively, cells can be immediately used for nuclei preparation without freezing.

All subsequent steps are carried out on ice or in a cold room at 4 °C.

5. Cells are washed three times in 0.9 mL cold buffer A. After each wash, cells are pelleted for 2 min with 16,000 × *g* at 4 °C, and the supernatant is discarded.
6. After the last washing step cells are suspended in 350 µL buffer A and ~500 µL cold glass beads (diameter 0.75–1 mm) are added (*see* **Note 9**).
7. Cell disruption takes place in a Vibrax Shaker at 4 °C at 2200 rpm (maximum speed) for 10 min.
8. For removal of glass beads and collection of cell lysate the microtubes are inverted and the tube tips pierced with a hot syringe needle. The tubes are placed with the tip to the bottom in a 15 mL reaction tube before the microtube lid is pierced with a hot syringe needle. The cell lysate is recovered from the microtube in the tip of the 15 mL reaction tube after centrifugation for 2 min with 130 × *g* at 4 °C in a benchtop centrifuge.
9. Cell lysate is transferred to a new 1.5 mL microtube and crude nuclei are pelleted for 1 min with 16,000 × *g* at 4 °C. The supernatant is discarded, and the nuclei are washed with 900 µL cold buffer A and pelleted for 1 min with 16,000 × *g* at 4 °C.
10. After removal of the supernatant the crude nuclei suspension can be frozen in liquid nitrogen and stored at −80 °C. Alternatively, nuclei can be immediately used for ChEC without freezing.

3.2 Chromatin Endogenous Cleavage (ChEC), DNA Extraction, and Restriction Enzyme Digest

1. Cell lysate is resuspended in 450 µL cold buffer Ag (*see* **Note 10**) and preincubated at 30 °C for 2 min under shaking with 750 rpm in a thermomixer.
2. After thoroughly mixing the cell lysate using a vortex shaker (*see* **Note 11**) two times 80 µL are taken as controls (0 min ChEC, no MNase activation). The control samples are incubated together with the ChEC sample at 30 °C until the last aliquot of the time course is withdrawn.
3. For the ChEC reaction 0.1 M $CaCl_2$ is added to the remainder of the cell lysate to a final concentration of 2 mM. Cell suspensions are incubated at 30 °C while shaking at 750 rpm in a thermomixer.
4. At different time points 80 µL aliquots are taken (*see* **Notes 11** and **12**) and added to microtubes containing 100 µL IRN

buffer to inactivate the MNase. After thoroughly mixing the samples with IRN they can be kept at room temperature until the end of the time course.

5. After the last aliquot of the time course is withdrawn, 100 µL of IRN buffer are added to the control samples (*see* **step 2**).
6. 10 µL of 10 % SDS and 2 µL Proteinase K (20 mg/mL) are added to the samples, and the mixture is incubated for 1 h at 56 °C.
7. Formaldehyde cross-links are reverted upon overnight incubation of the samples at 65 °C (*see* **Note 13**).
8. An equal volume of phenol/chloroform/isoamyl alcohol (1:1:1) mixture is added to the samples followed by thoroughly mixing (~20 s). Centrifugation for at least 1 min with 16,000 × *g* at room temperature leads to phase separation.
9. 120 µL of aqueous phase (upper phase, containing the DNA) are transferred to a new microtube.
10. An equal volume of IRN buffer is added (120 µL) and nucleic acids are precipitated by the addition of 2.5 volumes of ice-cold ethanol p.a. After mixing, samples are kept for 15 min at –20 °C.
11. Nucleic acids are pelleted by centrifugation for at least 20 min with 16,000 × *g* at 4 °C, and the supernatant is discarded (a water jet pump can be used for this purpose).
12. An optional washing step with 150 µL 70 % ethanol p.a. can be carried out to remove salt and residual phenol/chloroform/isoamyl alcohol, but is usually unnecessary.
13. Samples are air-dried for approximately 30 min at room temperature (*see* **Note 14**).
14. DNA is resuspended in 40 µL RNase A-TE (0.05 mg/mL) and incubated for 1 h at 37 °C.
15. 12.5 µL of RNase A treated DNA are digested with an appropriate restriction enzyme in a total volume of 20 µL overnight according to the manufacturer's recommendations.

3.3 Agarose Gel Electrophoresis and Southern Blot

1. 10× loading buffer is added to the samples, which are separated in a 1 % TBE agarose gel (250 mL, 15 × 20 cm, with nucleic acid stain).
2. Electrophoresis is carried out at 6 V/cm until the bromophenol blue band (migrating at around 500 bp in a 1 % agarose gel) is approximately 1–2 cm from the lower edge of the gel (run time approximately 6 h).
3. A digital fluorescent image of the agarose gel is recorded using the FLA-3000 imaging system (Figs. 1a and 2a).
4. Gels are equilibrated two times in approximately 500 mL denaturing solution for 15 min (*see* **Note 15**).

5. Gels are washed two times in approximately 500 mL blotting solution for 15 min.
6. The DNA in the agarose gel is subjected to overnight upward capillary transfer to a positively charged membrane according to the standard Southern blot procedure [13].
7. After the transfer, blotting membranes are dried and DNA is cross-linked to the membrane using a Fluo-Link DNA Fixation and Visualization Transilluminator (254 nm, 0.3 J/cm^2).

3.4 Probe Preparation, Membrane Hybridization, Washing, and Membrane Stripping

1. The membrane is transferred to a hybridization tube and incubated with 50 mL prewarmed hybridization buffer for at least 1 h at 65 °C in a hybridization oven (*see* **Note 16**).
2. Radioactively labeled probes are prepared according to the manufacturer's recommendations using the RadPrime DNA labeling System with [α-^{32}P]dATP.
3. The labeled probes are purified from unincorporated [α-^{32}P] dATP using ProbeQuant™ G-50 Micro Columns following the instructions of the manufacturer.
4. Salmon sperm DNA is added to the probe (to yield a final concentration of 0.1 μg/mL in the hybridization solution).
5. The probe is incubated at 95 °C in a thermomixer for 10 min and then quickly chilled on ice.
6. The hybridization buffer is discarded from the hybridization tube and replaced by 15 mL fresh preheated hybridization buffer to which the labeled probe is added. Hybridization is carried out overnight at 65 °C in a hybridization oven.
7. The hybridization solution is discarded from the membrane and the membrane is washed once with 30 mL rinse buffer (a few rotations in the hybridization oven).
8. Rinse buffer is discarded and the membrane is washed twice at 65 °C with 30 mL wash buffer 1 for 15 min. After each wash step the old buffer is discarded.
9. The same procedure is repeated with wash buffers 2 and 3.
10. The membrane is dried, put in a plastic bag, and exposed to a phosphor imaging screen.
11. Radioactive signals on the blot are visualized by a FLA-3000 imaging system (Figs. 1b and 2b).
12. For stripping of the probe, 100 mL of (boiling) hot stripping solution are added to the membrane in a hybridization tube which is then incubated for 15 min at 80 °C in a hybridization oven. This procedure is repeated three times. The successful probe removal can be monitored by checking the membrane before and after stripping with a Geiger–Müller counter. The stripped membrane can then be subjected to another round of hybridization with a different probe (*see* **Note 17**).

3.5 Quantification of Degradation

1. Quantification and analysis of ChEC signals are performed with the Multi Gauge v3.0 software.
2. In the "Measure" mode the "Quant" module allows to measure the signal intensity at selected regions of interest on the membrane.
3. To measure the degradation in ChEC experiments the regions on the membrane containing the full-length fragments are selected in each individual lane (*see* **Note 18**). One additional region is selected where no radioactive signal is detectable. The value obtained for this region is the "background" which has to be subtracted from the values measured for the full-length fragments.
4. The obtained data can be exported to MS Excel. To determine the degradation of the full-length fragment, the values obtained for the full-length fragment in the samples of the different ChEC time points are divided by the value obtained for the control samples (0 min ChEC, no MNase activation, *see* Subheading 2.2). The same analysis is carried out for signals obtained with different radioactively labeled probes detecting different genomic loci (Figs. 1c and 2c). The values obtained for a genomic region of interest to which the MNase fusion protein binds can be normalized to the values obtained for a genomic region to which no binding of the MNase fusion protein is observed (Fig. 1c, graph "after normalization"). For genome-wide binding factors a normalization to values obtained for the degradation of a plasmid added prior to ChEC to the reaction mixture is recommended (*see* **Note 12**).

4 Notes

1. Yeast extract (1 % w/v), yeast peptone (2 % w/v), and dextrose (2 % w/v) are used as growth medium.
2. Adenine hemisulfate should be added to the medium (100 mg/L) if the yeast strain is auxotrophic for adenine. This is especially important for strains carrying the *ade2* mutation because of the accumulation of a toxic intermediate.
3. The glycine solution should be autoclaved for long-term storage.
4. EDTA concentration can be increased up to 20 mM without affecting the experimental outcome. Higher EDTA concentrations are especially useful for cells that may have higher intracellular Ca^{2+} concentrations (e.g., stationary yeast cells) to avoid premature MNase activation.
5. For the analysis of chromatin composition with transition to stationary phase we recommend to take samples at the diauxic

shift (app. OD_{600} 5–7, app. 6–9 h since OD_{600} 0.5), during the post-diauxic growth phase (app. OD_{600} 10–13, app. 48 h after OD_{600} 0.5), and in stationary phase (app. OD_{600} 15–20, app. 144 h after OD_{600} 0.5).

6. To achieve a similar cell number in all samples the volume taken from the growing culture has to be adjusted. With transition to stationary phase the OD_{600} values do no longer correlate with the cell number. Instead the observed increase in OD_{600} might reflect growth of individual cells and changes in the morphology of the cell wall. The latter correlates with a lower efficiency of cell lysis during nuclei preparation. Thus, the following aliquot volumes are recommended to isolate similar amounts of genomic DNA from the different samples: mid-exponential phase 50 mL, all other samples starting from diauxic shift 5 mL.
7. To obtain similar formaldehyde cross-linking conditions for the different samples, cell aliquots with lower volume (*see* **Note 6**) are filled up with prewarmed YPD to a total volume of 50 mL matching the volume of the sample taken at mid-exponential growth phase.
8. Cross-linking temperature (and time) is crucial for ChEC efficiency and can be a reason for variation in DNA degradation by MNase. At higher temperatures protein-protein and protein-DNA cross-linking is stronger than at lower temperature. If the temperature is too low, the cross-link can be incomplete thus leading to free MNase fusion proteins and unspecific degradation. If the temperature is too high MNase activity can be compromised.
9. The volume of beads should be chosen so that the beads are still covered by a thin liquid layer of buffer A.
10. The reaction volume should be at least 50 μL higher than the total volume of aliquots taken during the ChEC time course.
11. Suspension has to be mixed vigorously to counteract nuclei sedimentation.
12. ChEC Time points have to be chosen according to the abundance of the MNase fusion protein. High-abundance proteins (e.g., histones) need shorter time intervals to avoid unspecific cleavage events, while low-abundance proteins (e.g., Pol I initiation factors) may be incubated for longer time intervals to maximize specific cleavage. MNase fusion proteins are cutting DNA near their binding sites and may be released from the DNA during ChEC. Released MNase fusion proteins act as free MNase. This might be the case for later time points of ChEC with cells expressing Hho1-MNase (Fig. 2b). To monitor unspecific cleavage mediated by free MNase fusion proteins, naked DNA (plasmid DNA) can be added to the cell

lysate. If a probe hybridizing to this plasmid DNA is used in the subsequent Southern blot analysis unspecific degradation can be detected. Samples in wich unspecific degradation is observed should be excluded from the quantitative analysis (Subheading 3.5).

13. Cross-link reversal is crucial for the experiment; thus incubation at 65 °C should last for at least 10 h.
14. To reduce time samples can be dried for app. 10 min at 60 °C in a thermomixer (constant control is recommended to avoid overdrying).
15. For efficient transfer of fragments larger than 15 kb, or increased hybridization efficiency of the probe with circular DNA the agarose gels should be incubated with 0.25 M HCl for 20 min at room temperature prior to adding the denaturing solution.
16. Up to four membranes can be stacked into one hybridization tube. The volume of the hybridization solution should be adjusted to cover all the membranes. Nylon meshes should be placed between the membranes to ensure access of the probe to the membrane surface.
17. After stripping, each blot can be hybridized several times without significantly affecting hybridization efficiency. As described in the section for quantification of degradation (Subheading 2.5) probing the blot to visualize a fragment of a genomic locus where the MNase fusion protein is not binding, can be used to determine the DNA load in each lane. This is an important information to interpret properly the degradation of a fragment of a locus to which the MNase fusion protein binds. If single-copy loci and multicopy loci are analyzed, the single-copy locus should be probed first. Probe removal is not complete which would result in a strong background if the multicopy locus probe is used first for hybridization. If MNase fusion proteins bind genome-wide plasmid DNA added to the nuclei before ChEC (*see* **Note 12**) can be used as a loading control.
18. Since the program calculates the pixel density within the selected region all selected regions should have same size.

Acknowledgements

We thank the members of the department of Biochemie III for constant discussion and support. This work has been supported by a grant of the DFG in the context of the SFB960 “Ribosome formation: principles of RNP biogenesis and control of their function” to H.T., P.M., and J.G.

References

1. Collas P (2010) The current state of chromatin immunoprecipitation. Mol Biotechnol 45: 87–100
2. Schmid M, Durussel T, Laemmli UK (2004) ChIC and ChEC: genomic mapping of chromatin proteins. Mol Cell 16:147–157
3. Schmid M, Arib G, Laemmli C et al (2006) Nup-PI: the nucleopore-promoter interaction of genes in yeast. Mol Cell 21:379–391
4. Ohmayer U, Perez-Fernandez J, Hierlmeier T et al (2012) Local tertiary structure probing of ribonucleoprotein particles by nuclease fusion proteins. PLoS One 7:e42449
5. Mao C, Brown CR, Griesenbeck J, Boeger H (2011) Occlusion of regulatory sequences by promoter nucleosomes in vivo. PLoS One 6: e17521
6. Merz K, Hondele M, Goetze H et al (2008) Actively transcribed rRNA genes in S. cerevisiae are organized in a specialized chromatin associated with the high-mobility group protein Hmo1 and are largely devoid of histone molecules. Genes Dev 22:1190–1204
7. Nogi Y, Yano R, Nomura M (1991) Synthesis of large rRNAs by RNA polymerase II in mutants of Saccharomyces cerevisiae defective in RNA polymerase I. Proc Natl Acad Sci U S A 88:3962–3966
8. Schäfer G, McEvoy CRE, Patterton H-G (2008) The Saccharomyces cerevisiae linker histone Hho1p is essential for chromatin compaction in stationary phase and is displaced by transcription. Proc Natl Acad Sci U S A 105:14838–14843
9. Claypool JA, French SL, Johzuka K et al (2004) Tor pathway regulates Rrn3p-dependent recruitment of yeast RNA polymerase I to the promoter but does not participate in alteration of the number of active genes. Mol Biol Cell 15:946–956
10. Dammann R, Lucchini R, Koller T, Sogo JM (1993) Chromatin structures and transcription of rDNA in yeast Saccharomyces cerevisiae. Nucleic Acids Res 21:2331–2338
11. French SL, Osheim YN, Cioci F et al (2003) In exponentially growing Saccharomyces cerevisiae cells, rRNA synthesis is determined by the summed RNA polymerase I loading rate rather than by the number of active genes. Mol Cell Biol 23:1558–1568
12. Wittner M, Hamperl S, Stöckl U et al (2011) Establishment and maintenance of alternative chromatin states at a multicopy gene locus. Cell 145:543–554
13. Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98: 503–517
14. Wu C (1980) The 5′ ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature 286:854–860
15. Goetze H, Wittner M, Hamperl S et al (2010) Alternative chromatin structures of the 35S rRNA genes in Saccharomyces cerevisiae provide a molecular basis for the selective recruitment of RNA polymerases I and II. Mol Cell Biol 30:2028–2045

Chapter 15

Selection and Validation of Spacer Sequences for CRISPR-Cas9 Genome Editing and Transcription Regulation in Bacteria

Frédéric Grenier, Jean-François Lucier, and Sébastien Rodrigue

Abstract

RNA-guided Cas9 nucleases derived from clustered regularly interspaced short palindromic repeat (CRISPR)-Cas systems have recently been adapted as sequence-programmable tools for various purposes such as genome editing and transcriptional regulation. A critical aspect of the system is the selection and validation of spacer sequences that allow precise targeting of the guide RNA-Cas9 complex. We describe a procedure involving computational and experimental steps to identify and test potentially interesting spacer sequences in bacterial genomes.

Key words CRISPR, Cas9, gRNA, Genome editing, Transcription, Repression

1 Introduction

Precise genome editing and gene transcription control are fundamental in the study of biological systems. While different approaches have been developed for these purposes [1], clustered regularly interspaced short palindromic repeat (CRISPR)-Cas systems are increasingly popular. In nature, the CRISPR-Cas system plays a role in bacterial adaptative immunity by degrading bacteriophage or conjugative plasmid DNA through an RNA-guided DNA nuclease complex [2, 3]. The activity of type II CRISPR-Cas can be recapitulated using only the Cas9 protein (*see* Fig. 1a) and a short guide RNA (gRNA, *see* Fig. 1c). The relative simplicity and ease of use this system have led to the development of a variety of applications in several organisms including both prokaryotes and eukaryotes [4–7]. The Cas9-gRNA complex specificity depends on direct hybridization between a 20-nucleotide sequence (referred to as the spacer, *see* Fig. 1c) located at the 5′ extremity of a guide RNA (gRNA), and the target DNA (the protospacer, *see* Fig. 1d). Another motif, the Protospacer Adjacent Motif (PAM), is needed

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_15,

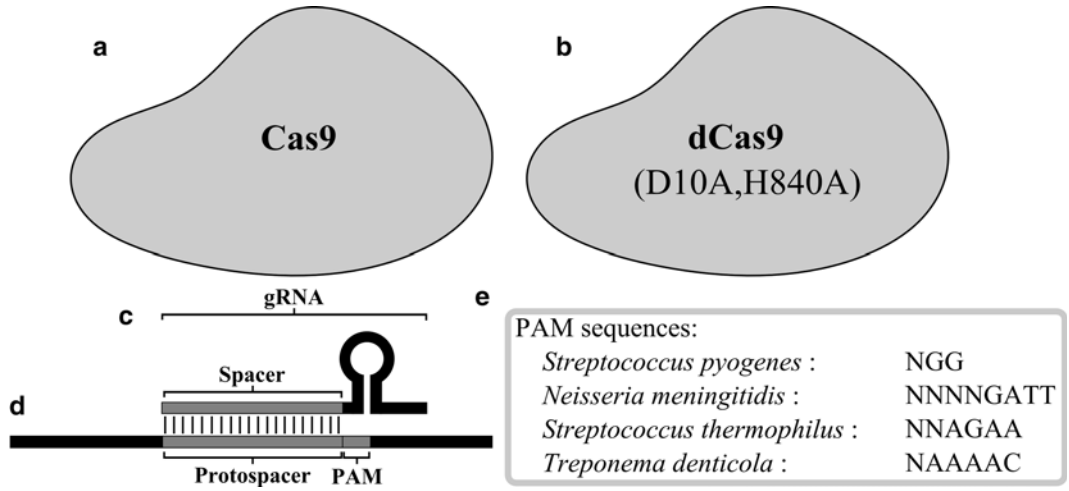


Fig. 1 Representation of Cas9-gRNA system elements. Endonuclease Cas9 (**a**), catalytically dead Cas9 (dCas9) (**b**), gRNA containing its spacer and the scaffold allowing Cas9 recognition (**c**), DNA containing the protospacer and the protospacer adjacent motif (PAM) (**d**), and different PAM sequences (**e**)

immediately 3′ of the protospacer for recognition by the Cas9-gRNA (*see* Fig. 1d) [8, 9]. PAM differs between Cas9 orthologs but the most commonly employed version from *Streptococcus pyogenes* consists of the trinucleotide 5′-NGG-3′ (*see* Fig. 1e) [10]. A key feature of the Cas9-gRNA system is that the spacer sequence can be exchanged to match virtually any protospacer (appropriately located next to a PAM), thus offering the possibility to program target specificity.

The Cas9 protein contains two nuclease domains, RuvC and HNH, each cleaving a strand of DNA upstream of the PAM [9, 11, 12]. Point mutations were introduced to abolish the activity of these domains, either independently to obtain "nickases" or simultaneously to generate a catalytically dead mutant (dCas9, *see* Fig. 1b) [9, 13–15]. The dCas9-gRNA complex can still recognize its target sequence but remains associated to DNA, which can be exploited to repress transcription by blocking promoter regions. Several protein domains have also been fused to the N- or C-terminus of dCas9, further expanding the list of possible applications with sequence-targeted transcription activation and visualization of specific loci [13, 16, 17].

In this chapter, we describe a bioinformatics tool allowing the selection of protospacer sequences, and experimental procedures to validate these sequences for Cas9-gRNA mediated endonuclease activity and dCas9-gRNA transcription repression in bacteria.

2 Materials

2.1 Selection of Spacer Sequences

1. Computer with browser and Internet access.

2.2 Construction of the gRNA Coding Plasmid

1. LB broth, sterilized by autoclave.
2. LB agar solid medium with appropriate antibiotics.
3. Ampicillin (50 mg/ml in water, 0.22 μm filter sterilized).
4. Tetracycline hydrochloride (15 mg/ml in ethanol 50 % v/v).
5. Shaking incubator.
6. *Escherichia coli* laboratory cloning strain.
7. Oligonucleotides for PCR amplification, available from several sources.
8. PCR amplification kits, available from many suppliers.
9. End-repair mix: T4 DNA Polymerase and T4 Polynucleotide Kinase supplied in 100 mM KCl, 10 mM Tris–HCl, 0.1 mM EDTA, 1 mM DTT, 0.1 % Triton X-100, 50 % glycerol, pH 7.4.
10. 10× stock solution of end-repair buffer: 1 M Tris–HCl, 500 mM NaCl, 100 mM $MgCl_2$, 50 mM DTT, 0.25 % Triton-X 100, pH 7.5.
11. Stock solution of 1 mM dNTPs.
12. T4 DNA ligase supplied in 10 mM Tris–HCl, 50 mM KCl, 1 mM DTT, 0.1 mM EDTA, 50 % glycerol, pH 7.4.
13. 2× stock solution of rapid ligation buffer: 132 mM Tris–HCl, 20 mM $MgCl_2$, 2 mM DTT, 2 mM ATP, 15 % PEG 6000, pH 7.6.
14. DpnI supplied in 10 mM Tris–HCl, 300 mM NaCl, 1 mM DTT, 0.1 mM EDTA, 500 μg/ml BSA, 50 % glycerol, pH 7.4.
15. 10× stock solution of digestion buffer: 500 mM potassium acetate, 200 mM Tris-acetate, 100 mM magnesium acetate, 1 mg/ml BSA, pH 7.9.
16. pFG001 (Addgene #62816) (*see* Fig. 3 and Table 1).
17. pFG018 (Addgene #62817) (*see* Fig. 3 and Table 1).
18. Thermal cycler.

2.3 Spacer Validation for Cas9-gRNA-Mediated Endonuclease Activity

2.3.1 Constitutive Endonuclease Activity

1. LB broth, sterilized by autoclave.
2. LB agar solid medium with appropriate antibiotics.
3. Ampicillin (50 mg/ml in water, 0.22 μm filter sterilized).
4. Kanamycin (50 mg/ml, 0.22 μm filter sterilized).
5. Chloramphenicol (34 mg/ml in ethanol 100 %).
6. Shaking incubator.

Table 1
Plasmid- and strain-associated antibiotic resistances

Name	Antibiotic resistance
pCMK	Kanamycin
pSIM7	Chloramphenicol
pFG001	Ampicillin
pFG018	Tetracycline hydrochloride
pdCas9-bacteria	Chloramphenicol
Escherichia coli BW25113-GFP-KanR	Kanamycin

7. Spectrophotometer.
8. *E. coli* laboratory strain.
9. pSIM7 (available from Donald Court, US-National Cancer Institute-Frederick) (*see* Table 1).
10. pCMK (Addgene #62818) (*see* Table 1).
11. pFG001 (Addgene #62816) (*see* Fig. 3 and Table 1).

2.3.2 Inducible Endonuclease Activity

1. LB broth, sterilized by autoclave.
2. LB agar solid medium with appropriate antibiotics.
3. Kanamycin (50 mg/ml, 0.22 μm filter sterilized).
4. Chloramphenicol (34 mg/ml in ethanol 100 %).
5. Tetracycline hydrochloride (15 mg/ml in ethanol 50 %).
6. L-arabinose (20 %, 0.22 μm filter sterilized).
7. D-glucose (20 %, 0.22 μm filter sterilized).
8. Shaking incubator.
9. Spectrophotometer.
10. *E. coli* laboratory strain.
11. pSIM7 (available from Donald Court, US-National Cancer Institute-Frederick) (*see* Table 1).
12. pCMK (Addgene #62818) (*see* Table 1).
13. pFG018 (Addgene #62817) (*see* Fig. 3 and Table 1).

2.4 Spacer Validation for Cas9-gRNA-Mediated Transcription Repression

1. LB broth, sterilized by autoclave.
2. LB agar solid medium with appropriate antibiotics.
3. Ampicillin (50 mg/ml in water, 0.22 μm filter sterilized).
4. Kanamycin (50 mg/ml, 0.22 μm filter sterilized).
5. Chloramphenicol (34 mg/ml in ethanol 100 %).
6. Spectrophotometer.

7. Shaking incubator.
8. *E. coli* BW25113-GFP-KanR (*see* Table 1).
9. pFG001 (Addgene #62816) (*see* Fig. 3 and Table 1).
10. pdCas9-bacteria (Addgene #44249) (*see* Table 1).
11. Plate reader with absorbance and fluorescence capacity.

3 Methods

Many target sites can be available for Cas9-gRNA recognition but selecting the sequence offering the highest probability of success can be challenging. Bioinformatics tool can help choosing the most specific sequence with the lowest chance of non-specific targeting and avoid gRNA secondary structures that could compromise its activity (*see* Subheading 3.1). Still, empirical variations between apparently equivalent gRNA are often observed, and experimental validation can help determine the most efficient sequence. This implies the cloning of selected spacers in plasmids designed for gRNA expression (*see* Subheading 3.2), which are used in conjunction with a Cas9 encoding plasmid to target either a plasmid or the genome that contains the targets for endonuclease activity assay (*see* Subheading 3.3) or transcription repression assay (*see* Subheading 3.4).

3.1 Selection of Spacer Sequences

1. Browse to the following website: http://bioinfo.ccs.usherbrooke.ca/cgi-bin/CrispyCrunch/index.pl (*see* **Note 1**).
2. Paste the sequence of interest that should be targeted by the gRNA in the "Target DNA Sequence" box (*see* Fig. 2a).
3. Select the appropriate "Off-target blast database" and click on "submit query" (*see* Fig. 2a).
4. The returned table lists possible spacer sequences and their characteristics (*see* Fig. 2b).
5. From the list, select interesting spacer sequences that can be empirically tested by following the sections below.

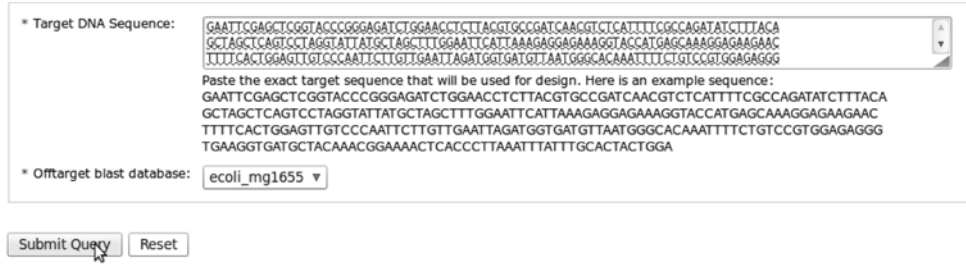
3.2 Construction of gRNA Coding Plasmids

Integrate the selected spacers in plasmids designed for gRNA expression (pFG001 for constitutive expression or pFG018 for arabinose inducible expression).

1. This will be done by PCR using pFG001 or pFG018 as template, a first primer that contains the desired spacer (*see* Fig. 3, primer 3) and a second constant primer (*see* Fig. 3, primer 1 or 5 for constitutive or inducible gRNA expression, respectively).

a

Crispy CRISPR design parameters



b

Crispy CRISPR predictions

Note: If you need help interpreting this result table, see information below prediction table or click here.

Nbr	Spacer sequence	Target sequence	Localisation bin	Position	Internal DG	Hybridisation DG	Offtargets
1	GGAGATCTGGAACCTCTTAC	GTAAGAGGTTCCAGATCTCC	1	20	-16.458	-25.8709	
2	GGGAGATCTGGAACCTCTTA	TAAGAGGTTCCAGATCTCCC	1	19	-17.27	-26.7596	
3	CTTACGTGCCGATCAACGTC	GACGTTGATCGGCACGTAAG	1	35	-18.392	-29.0146	
4	GATCTCCCGGGTACCGAGCT	AGCTCGGTACCCGGGAGATC	1	7	-20.047	-30.0581	
5	GATATCTTTACAGCTAGCTC	GAGCTAGCTGTAAAGATATC	1	67	-20.964	-24.1551	
6	CTTTAATGAATTCCAAAGCT	AGCTTTGGAATTCATTAAAG	2	105	-16.526	-23.6468	
7	TTGCTCATGGTACCTTTCTC	GAGAAAGGTACCATGAGCAA	2	127	-17.011	-26.2393	
8	TTCTCCTCTTTAATGAATTC	GAATTCATTAAAGAGGAGAA	2	112	-17.222	-22.8385	
9	GTGAAAAGTTCTTCTCCTTT	AAAGGAGAAGAACTTTTCAC	2	145	-18.139	-24.3852	
10	AGGTATTATGCTAGCTTTGG	CCAAAGCTAGCATAATACCT	2	93	-19.501	-24.9852	
11	TGAGCAAAGGAGAAGAACTT	AAGTTCTTCTCCTTTGCTCA	2	140	-20.628	-25.9559	
12	AAGCTAGCATAATACCTAGG	CCTAGGTATTATGCTAGCTT	2	90	-20.324	-24.3769	
13	AGGACTGAGCTAGCTGTAAA	TTTACAGCTAGCTCAGTCCT	2	73	-21.522	-26.34	
14	TTAACATCACCATCTAATTC	GAATTAGATGGTGATGTTAA	3	187	-16.683	-22.9258	
15	ATTCTTGTTGAATTAGATGG	CCATCTAATTCAACAAGAAT	3	178	-17.102	-23.1822	
16	ATTAACATCACCATCTAATT	AATTAGATGGTGATGTTAAT	3	188	-17.367	-22.3941	
17	AATTCTTGTTGAATTAGATG	CATCTAATTCAACAAGAATT	3	177	-18.525	-22.3589	
18	TCTAATTCAACAAGAATTGG	CCAATTCTTGTTGAATTAGA	3	175	-19.457	-23.5619	
19	CGGACAGAAAATTTGTGCCC	GGGCACAAATTTTCTGTCCG	3	208	-20.216	-27.9491	
20	CTCCACGGACAGAAAATTTG	CAAATTTTCTGTCCGTGGAG	3	213	-20.026	-26.4424	
21	TCTCCACGGACAGAAAATTT	AAATTTTCTGTCCGTGGAGA	3	214	-21.22	-26.3385	
22	TGGAGAGGGTGAAGGTGATG	CATCACCTTCACCCTCTCCA	4	228	-15.575	-27.9416	
23	TTAAATTTATTTGCACTACT	AGTAGTGCAAATAAATTTAA	4	269	-16.736	-21.8836	
24	TTTGTAGCATCACCTTCACC	GGTGAAGGTGATGCTACAAA	4	235	-17.011	-26.5626	
25	TCACCCTCTCCACGGACAGA	TCTGTCCGTGGAGAGGGTGA	4	220	-20.753	-29.831	
26	ATCAACGTCTCATTTTCGCC	GGCGAAAATGAGACGTTGAT	1	46	-17.849	-27.4221	NC_000913.3: (16/20) |||||||.||||.||||..
27	TAAATTTATTTGCACTACTG	CAGTAGTGCAAATAAATTTA	4	270	-16.803	-23.0408	NC_000913.3: (17/20) |||||.||.|||||||||.
28	GTAGTGCAAATAAATTTAAG	CTTAAATTTATTTGCACTAC	4	268	-16.152	-22.0807	NC_000913.3: (16/20) |||||||||||.||||...

Fig. 2 CRISPR design tool web interface. Search screen (**a**) and result table (**b**) generated by the Crispy CRISPR design tool. The result table contains a list of in-silico spacer sequences predictions ranked by likelihood of being functional. The results are sorted based on these parameters: (1) predictions with the least to the most number of off-targets, (2) the localization in the submitted DNA region to be targeted, (3) the folding of the spacer sequence to favor sequence with the lowest potential of generating intra molecular secondary structures, and finally (4) hybridization energy between the spacer and the targeted sequence to favor spacer sequence with high affinity for its target sequence. These last two filters where computed using *mfold* software [22]

2. End-repair to ensure that the ends of the PCR product are blunt and 5′ phosphorylated to allow subsequent ligation (*see* **Note 2**):

   (a) Set up a 25 μl end-repair reaction containing up to 1.0 μg of purified PCR product, 2.5 μl 10× end-repair buffer, 2.5 μl 1 mM dNTPs, 2 μl (20 U T4 Polynucleotide Kinase, 6 U T4 DNA Polymerase) end-repair enzyme mix, and complete with sterile water to a volume of 25 μl.

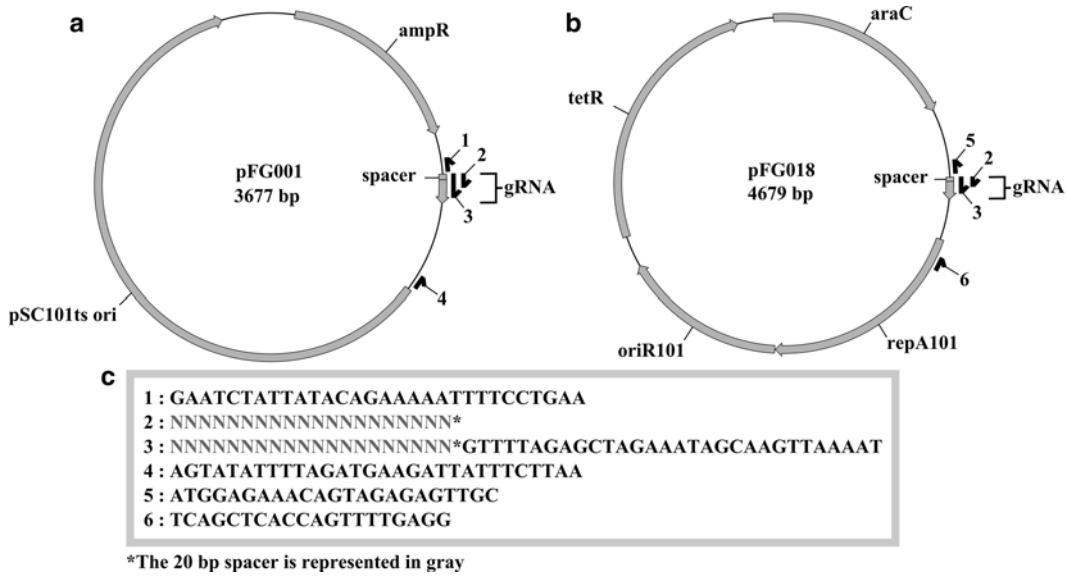


Fig. 3 Representation of gRNA expression plasmids used in this protocol. pFG001 (**a**), pFG018 (**b**), and primer sequence (5′–3′) (**c**)

(b) Incubate at room temperature for 30 min.

3. Circularize the PCR product:
   (a) Set up a 20 μl ligation reaction containing 20–200 ng of purified end-repaired PCR product, 10 μl 2× rapid ligation buffer, 1 μl (600 U) T4 DNA ligase, and sterile water to a final volume of 20 μl.
   (b) Incubate at room temperature for 20 min.
4. Eliminate the PCR template:
   (a) Set up a 10 μl digestion reaction containing up to 1.0 μg of purified circularized product, 1 μl of digestion buffer, 1 μl (20 U) DpnI (*see* **Note 3**), and sterile water to a final volume of 10 μl.
5. Transform the resulting plasmid in an *Escherichia coli* cloning strain.
6. Spread the transformed cells on LB agar solid medium with either ampicillin (50 μg/ml) for pFG001 or tetracycline hydrochloride (15 μg/ml) for pFG018 (*see* Table 1).
7. To confirm the integration of the spacer, screen the transformants by colony PCR using a primer identical to the spacer (*see* Fig. 3, primer 2) and a constant primer (*see* Fig. 3, primer 4 or 6 for pFG001 or pFG018, respectively) (*see* **Note 4**). Plasmids isolated from positive clones can be subjected to direct sequencing to further validate that the correct gRNA will be produced.

3.3 Spacer Validation for Cas9-gRNA-Mediated Endonuclease Activity

3.3.1 Constitutive Endonuclease Activity

A simple way to test the endonuclease activity of a selected gRNA sequence is to target a plasmid containing the corresponding protospacer along with the appropriate PAM motif (*see* **Note 5**). Here, a gRNA expressing plasmid (pFG001) is transformed into cells already containing two other plasmids, one coding for a Cas9 protein (pCMK) and another one containing the protospacer (pSIM7) to see if the targeted plasmid is efficiently lost. This general scheme can be applied to validate most gRNA provided that the target is not found in the *E. coli* genome by cloning the targeted protospacer and PAM sequences in an adequate plasmid (*see* **Note 6**).

1. Transform pFG001 plasmids in cells containing pCMK and pSIM7. A pFG001 plasmid with a test spacer (targeting a protospacer that is not present in host DNA) can be used as a negative control.
2. Spread the transformed cells on LB agar solid medium with ampicillin (50 μg/ml), kanamycin (50 μg/ml), and on LB agar with ampicillin (50 μg/ml), kanamycin (50 μg/ml), and chloramphenicol (34 μg/ml).
3. Incubate at 30 °C overnight.

Bacterial growth is expected on both media for the test spacer since it should not allow the Cas9-gRNA complex to cleave any site in the cell. Appropriate spacers should not yield growth on medium containing chloramphenicol, indicating that the targeted plasmid (pSIM7) was cleaved and lost (*see* Fig. 4a and Table 1).

3.3.2 Inducible Endonuclease Activity

For some applications, gRNA expression can be preferred under specific conditions or at particular times. For this reason, we have developed an inducible gRNA expression system that can be triggered by the presence of arabinose (*see* **Note 7**). The following experiment is similar to the previous, except that gRNA transcription is under control of an arabinose inducible promoter (*see* **Note 8**).

1. Transform pFG018 plasmids in cells containing pCMK and pSIM7 (*see* **Note 9**). A pFG018 plasmid with a test spacer (targeting a protospacer that is not present in the cells' DNA) can be used as a negative control.

a

pFG001-test or pFG001 → *E. coli* + pCMK + pSIM7

Medium	Growth	
	pFG001-test	pFG001
LB agar + ampicillin/kanamycin	+++	+++
LB agar + ampicillin/kanamycin/chloramphenicol	+++	-

b

pFG018-test or pFG018 → *E. coli* + pCMK + pSIM7

Medium	Growth	
	pFG018-test	pFG018
LB agar + tetracycline/kanamycin/chloramphenicol + glucose 5%	+++	+++
LB agar + tetracycline/kanamycin/chloramphenicol + arabinose 1%	+++	-

Fig. 4 Endonuclease activity assay. Tests using constitutive (**a**) or inducible (**b**) gRNA expression

2. Spread the transformed cells on LB agar solid medium with tetracycline hydrochloride (15 μg/ml), kanamycin (50 μg/ml), chloramphenicol (34 μg/ml), D-glucose 5 %, and on LB agar with tetracycline hydrochloride (15 μg/ml), kanamycin (50 μg/ml), chloramphenicol (34 μg/ml), and L-arabinose 1 %.
3. Incubate at 30 °C overnight.

Growth on the medium containing D-glucose but not on the medium containing L-arabinose indicates that the tested spacers are functional. Test spacers should yield growth on both medium (*see* Fig. 4b and Table 1).

3.4 Spacer Validation for Cas9-gRNA-Mediated Transcription Repression

Transcription repression potential of a dCas9-gRNA complex can be assessed by two approaches. First, the promoter of the gene of interest can be cloned upstream of a reporter for which the activity will be measured in presence or absence of the gRNA in cells expressing Cas9. Alternatively, if the targeted sequence is naturally found in the chromosome, reverse-transcriptase quantitative PCR (RT-qPCR) can directly be performed on the targeted gene. An experiment in which a GFP encoding gene and the targeted promoter are integrated in the genome of *E. coli* BW25113 [18] is described (*see* **Note 10**).

1. Integrate a cassette containing the genes coding for the GFP protein and the neomycin phosphotransferase (conferring kanamycin resistance) in the genome of *E. coli* BW25113 using λ-Red recombination system [19, 20].
2. Transform the resulting strain, *E. coli* BW25113-GFP-KanR, with pdCas9-bacteria, coding for catalytically inactive Cas9 (dCas9), and with pFG001 plasmids coding for gRNAs targeting different protospacers along the promoter and coding sequence (*see* Fig. 5b and Table 1).
3. Grow a culture of the transformants by inoculating single colonies into tubes containing 5 ml of LB broth with ampicillin (50 μg/ml), kanamycin (50 μg/ml), and chloramphenicol (34 μg/ml).
4. Incubate the cultures at 30 °C overnight.
5. Measure the emitted fluorescence (excitation = 485 nm, emission = 528 nm) with a 96-well plate reader, as well as the OD[600] for value normalization.

Different levels of fluorescence will be detected depending on the binding ability of the spacers and on their position along the promoter (*see* Fig. 5). Low fluorescence indicates a good repression associated with the tested spacer.

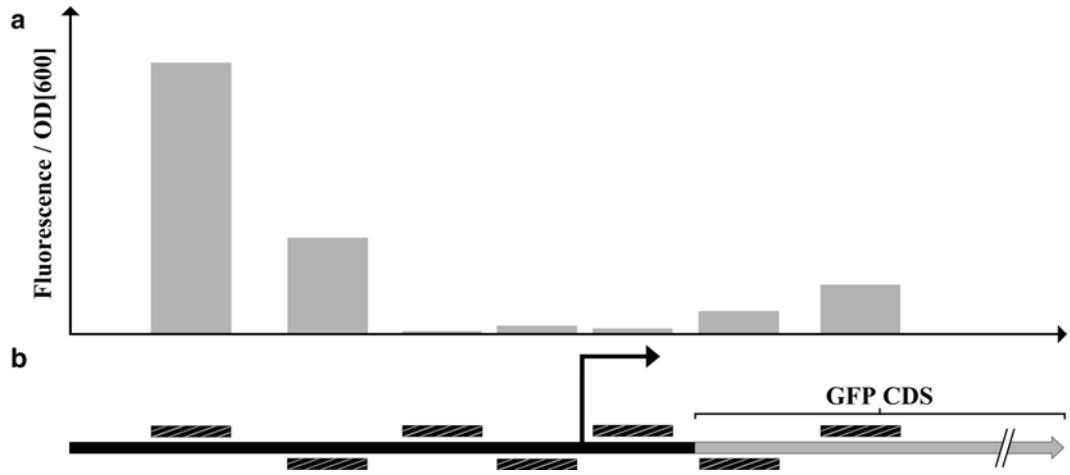


Fig. 5 Transcription repression assay: Normalized fluorescence of cells expressing spacers that target different protospacers on both strands of a promoter driving the expression of a GFP reporter gene (**a**). Schematic representation of the GFP promoter where the transcription start site is represented by an *arrow* and tested spacers are shown by *hatched bars* (**b**). As shown in this figure, the level of repression is generally higher when the dCas9-gRNA complex is targeted in the region surrounding the transcription start site

4 Notes

1. Alternative gRNA design tools are available online.
2. To obtain a product adequate for ligation, it can be either amplified with 5′ phosphorylated primers or end-repaired after amplification.
3. The restriction enzyme DpnI cleaves the sequence Gm^6A^TC. PCR products are not methylated and therefore will not be cleaved. Plasmid DNA obtained by in vivo amplification is generally methylated and will be cut by DpnI.
4. The screening is done to confirm that the plasmid contains the desired spacer and is not the initial plasmid used as the template for the PCR reaction.
5. Targeting a plasmid can reveal the presence of off-target cleavage sites in the genomic DNA. This phenomenon becomes apparent when cell death is observed even in the absence of the antibiotic used for selecting the targeted plasmid. Since cell death and low transformation efficiency can be hard to distinguish, the inducible gRNA expression system described in Subheading 3.3.2 can be advantageous. In this context, cells can be transformed to acquire all necessary plasmids and grown until the gRNA expression is triggered by addition of arabinose in the medium. Comparing the induced and control culture each in presence or absence of antibiotic selecting for the targeted plasmid should help identify off-target cleavage due

to the presence of cell death under induced gRNA expression independently of plasmid selection with the appropriate antibiotic.

6. The protospacer and the PAM can be cloned in a plasmid by PCR, using the same method described in Subheading 3.2.
7. Arabinose-inducible systems use the regulatory compounds of the arabinose operon. This operon can be regulated both positively and negatively. When arabinose is the main carbon source, it binds to the regulatory protein AraC and, with the cAMP-CRP complex, stimulates the expression of the catabolic genes downstream the P_{BAD} promoter. In the absence of arabinose, AraC causes the folding of the operon that represses the catabolic genes. Furthermore, high levels of glucose induce a decrease of cAMP level, thereby avoiding the formation of the cAMP-CRP complex which enhance the repression [21].
8. The transformation step can be separated from the activation of gRNA transcription, avoiding misleading results that could be obtained because of low transformation efficiency.
9. The cells must grow in 5 % glucose medium to maximize the catabolic repression of the arabinose-inducible P_{BAD} promoter by CRP and avoid unwanted gRNA expression.
10. The GFP open reading frame and its promoter could also be cloned in a plasmid instead of integrated in the chromosome but results obtained with a plasmid should be interpreted more carefully since high plasmid copy number could titer the dCas9-gRNA complex, leading to high fluorescence of the cells.

Acknowledgement

We are grateful to Donald L. Court (NCI-Frederick) for the generous gift of pSIM7 and to Dominick Matteau and Alain Lavigueur for critical reading of the manuscript. We thank the *Centre de calcul scientifique* of Université de Sherbrooke for computational resources and technical support. This work was supported by the *Natural Sciences and Engineering Research Council* of Canada (*NSERC*). S.R. holds a Chercheur-boursier Junior 1 award from the Fonds de recherche Québec-Santé.

References

1. Esvelt KM, Wang HH (2013) Genome-scale engineering for systems and synthetic biology. Mol Syst Biol 9:641. doi:10.1038/msb.2012.66
2. Marraffini L, Sontheimer EJ (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet 11:181–190. doi:10.1038/nrg2749

3. Horvath P, Barrangou R (2010) CRISPR/Cas, the immune system of bacteria and archaea. Science 327:167–170. doi:10.1126/science.1179555
4. Jiang W, Bikard D, Cox D et al (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31:233–239. doi:10.1038/nbt.2508
5. Mali P, Esvelt KM, Church GM (2013) Cas9 as a versatile tool for engineering biology. Nat Methods 10:957–963. doi:10.1038/nmeth.2649
6. Sander JD, Joung JK (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol 32:347–355. doi:10.1038/nbt.2842
7. Hsu PD, Lander ES, Zhang F (2014) Development and applications of CRISPR-Cas9 for genome engineering. Cell 157:1262–1278. doi:10.1016/j.cell.2014.05.010
8. Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) PNAS Plus: Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci 109:E2579–E2586. doi:10.1073/pnas.1208507109
9. Jinek M, Chylinski K, Fonfara I et al (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337:816–822
10. Esvelt KM, Mali P, Braff JL et al (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nat Methods 10:1116–1121. doi:10.1038/nmeth.2681
11. Anders C, Niewoehner O, Duerst A, Jinek M (2014) Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature 513:569–573. doi:10.1038/nature13579
12. Nishimasu H, Ran FA, Hsu PD et al (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell 156:935–949. doi:10.1016/j.cell.2014.02.001
13. Mali P, Aach J, Stranges PB et al (2013) Cas9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat Biotechnol 31:833–838. doi:10.1038/nbt.2675
14. Cong L, Ran FA, Cox D et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science 339:819–823. doi:10.1126/science.1231143
15. Qi LS, Larson MH, Gilbert L et al (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152:1173–1183. doi:10.1016/j.cell.2013.02.022
16. Bikard D, Jiang W, Samai P et al (2013) Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res 41:7429–7437. doi:10.1093/nar/gkt520
17. Tanenbaum ME, Gilbert L, Qi LS et al (2014) A versatile protein tagging system for signal amplification in single molecule imaging and gene regulation. Cell 159:635–646. doi:10.1016/j.cell.2014.09.039
18. Grenier F, Matteau D, Baby V, Rodrigue S (2014) Complete genome sequence of Escherichia coli BW25113. Genome Announc 2:e01038. doi:10.1128/genomeA.01038-14
19. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97:6640–6645. doi:10.1073/pnas.120163297
20. Datta S, Costantino N, Court DL (2006) A set of recombineering plasmids for gram-negative bacteria. Gene 379:109–115. doi:10.1016/j.gene.2006.04.018
21. Schleif R (2010) AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. FEMS Microbiol Rev 34:779–796. doi:10.1111/j.1574-6976.2010.00226.x
22. Mathews DH, Sabina J, Zuker M et al (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol 288:911–940

Chapter 16

Detection of Short-Range DNA Interactions in Mammalian Cells Using High-Resolution Circular Chromosome Conformation Capture Coupled to Deep Sequencing

Jean-François Millau and Luc Gaudreau

Abstract

DNA interactions shape the genome to physically and functionally connect regulatory elements to their target genes. Studying these interactions is crucial to understanding the molecular mechanisms that regulate gene expression. In this chapter, we present a protocol for high-resolution circular chromosome conformation capture coupled to deep sequencing. This methodology allows to investigate short-range DNA interactions (<100 kbp) and to obtain high-resolution DNA interaction maps of loci. It is a powerful tool to explore how regulatory elements and genes are connected together.

Key words 4C, Circular chromosome conformation capture, Short-range interactions

1 Introduction

An important part of gene regulation occurs through the action of distal regulatory elements that activate or repress promoters [1–3]. In order for these regulations to occur, DNA interactions need to be formed to physically and functionally connect regulatory elements to their target genes [1, 4, 5]. Studying DNA interactions is therefore crucial for understanding the precise molecular mechanisms behind gene regulation.

A decade ago, the chromosome conformation capture (3C) technique was developed to investigate DNA interactions [6]. This method is based on the principle that when two distant loci interact together, once their DNA is digested and then re-ligated, DNA fragments from one locus will be ligated to DNA fragments from the other locus because of their physical proximity. Subsequently, it is possible to quantify the interaction between the two loci by measuring the ligation products formed between a restriction site from one locus and a restriction site from the other locus using PCR. 3C is thus a "one-versus-one" technique;

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_16,

the interaction is measured using one restriction site versus another one. Consequently, each PCR reaction probes for one interaction defined by two restriction sites and one primer set. 3C is a powerful technique when one already suspects an interaction and wants to confirm it, as the restriction sites and primer set to use can easily be selected. However, it is less convenient for investigating new interactions because a tremendous number of PCR reactions are needed in order to map DNA interactions over megabases pair of DNA.

Circular chromosome conformation capture (4C) and the subsequent 4C coupled to deep sequencing (4C-seq) techniques were derived from 3C to overcome this disadvantage [7–9]. The 4C-seq method is based on the same principle, but takes advantage of circularizing DNA to amplify all the captured interacting sequences in one PCR reaction. It is thus a "one-versus-all" technique because it will measure and capture all the interactions occurring at one restriction site. Basically, a 4C-seq experiment can be divided into a few key steps. First, cells are fixed using formaldehyde to cross-link DNA interactions. Then, chromatin is digested using a first restriction enzyme and a first ligation round is performed to capture the interactions, as in 3C. Cross-links are then reversed and DNA is digested using a second restriction enzyme. This step helps to reduce the size of the captured fragments for efficient subsequent PCR. A second ligation step is then performed to circularize DNA, which allows the amplification of all captured fragments by a single PCR reaction. Using deep sequencing technology, the sequences of all captured and amplified interacting fragments are then obtained (*see* Fig. 1a).

An important feature of 4C-seq is the choice of the first cutter restriction enzyme because it directly dictates the resolution of the experiment. Usually, 4C-seq is performed using a six-cutter (a restriction enzyme whose restriction site is composed of six bases) to investigate long-range DNA interactions. Statistically, six-cutter sites are distributed every 4096 bp within the genome, which allows a measure of *cis*-interactions up to 5 Mbp from each side of a region of interest called the viewpoint. Nevertheless, in any 4C experiment, a large part of the measured interactions surround the viewpoint. These interactions are captured because of their linear proximity to the viewpoint and do not necessarily reflect functional interactions. In the case of a six-cutter, this region represents 200 kbp on each side of the viewpoint. Using a six-cutter as first cutter, it is thus difficult to measure functional short-range DNA interactions located within this 400 kbp "blind spot."

In this chapter we describe a protocol to perform high-resolution 4C-seq. This variant of the 4C-seq approach uses a four-cutter as first cutter. Because four-cutters cut DNA every

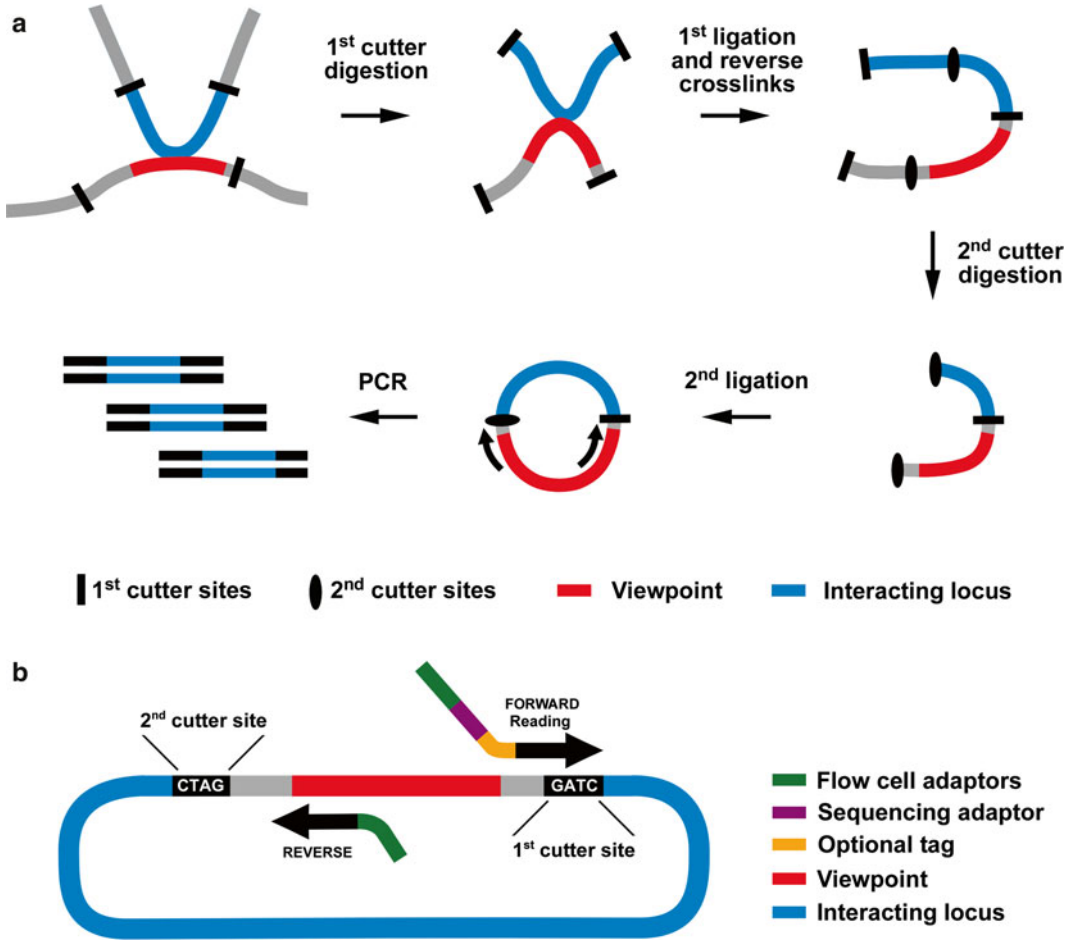


Fig. 1 (**a**) Overview of the circular chromosome conformation capture coupled to deep sequencing (4C-seq) technique. The interaction between a region of interest (*red*), also called viewpoint, and another locus (*blue*) is cross-linked in formaldehyde-fixed cells. The chromatin is then digested using a first cutter restriction enzyme. The resulting cohesive ends are then ligated, which captures the interaction between the region of interest and the interacting locus, and cross-links are reversed. To reduce the length of the captured interaction for optimal PCR amplification, the DNA is trimmed using a second cutter restriction enzyme. A second ligation is then performed to circularize the DNA followed by a PCR to amplify the interacting locus. PCR products are then ready to be deep sequenced. Note that for the sake of simplicity only one interaction was depicted, but the technique virtually captures all the interactions occurring at this viewpoint. (**b**) Design of the 4C-seq PCR primers. The forward reading primer overlaps and finishes on the first cutter restriction site (e.g., DpnII GATC). It is composed of a flow cell adaptor, a sequencing adaptor and a tag (optional). The Illumina sequencing reaction starts from the sequencing adaptor, hence the name "reading primer." The reverse primer should be within 100 bp of the second cutter (e.g., BfaI CTAG) site and also includes a flow cell adaptor

254 bp in average, it allows to measure interactions at high-resolution over 100 kbp from each side of the viewpoint with a "blind spot" of only 7 kbp. This technique is particularly well-suited for obtaining high-resolution maps of short-range DNA interactions.

2 Materials

2.1 Enzymes

1. First cutter restriction enzyme DpnII (NEB, R0543M).
2. Second cutter restriction enzyme BfaI (NEB, R0568L).
3. T4 DNA ligase (Roche, 10799009001).
4. Polymerase Expand Long Template PCR System (Roche, 11681842001).
5. Proteinase K (10 mg/ml).
6. RNase (10 mg/ml).

2.2 Solutions

1. Phenol-chloroform (buffer equilibrated pH 6–8).
2. 2.5 M Glycine.
3. 0.5 M EDTA.
4. 5 M NaCl.
5. 25 % Nonidet P 40 Equivalent (Sigma, 74385).
6. 20 % SDS.
7. 25 % Triton X-100.
8. 10 mM Tris–HCl pH 7.5.
9. 3 M Sodium acetate (NaAc) pH 5.2.
10. 70 and 100 % ethanol.
11. Glycogen (20 mg/ml) (Roche, 10901393001).
12. 25× cOmplete Protease Inhibitor Cocktail (Roche, 04693116001).
13. Formaldehyde 37 % free from acid (Merck, 1.03999.1000).
14. Methyl Green-Pyronin staining (EMS, 18710-02).
15. 100 mM ATP (Thermo Scientific, R0441).

2.3 Buffers

1. Lysis buffer: 50 mM Tris–HCl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5 % NP40, 1 % Triton-X100, 1× cOmplete Protease Inhibitor Cocktail (Roche).
2. 10× Ligation buffer: 660 mM Tris–HCl pH 7.5, 50 mM $MgCl_2$, 50 mM DTT, 10 mM ATP.

2.4 DNA Purification

1. QIAquick PCR Purification Kit (Qiagen, 28104).
2. Agencourt AMPure XP (Beckman Coulter, A63880).

2.5 Apparatus

1. PCR machine.
2. Eppendorf Thermomixer.
3. Nanodrop spectrophotometer.
4. 65 °C water bath.
5. 16 °C water bath.
6. Magnetic separation rack for 1.5 ml tubes.

3 Methods

3.1 Choosing the First and Second Cutters for Your Viewpoint

To perform high-resolution 4C-seq, two different four base cutter restriction enzymes are needed. One enzyme has to cut on the 5′ side of the viewpoint while the other one will cut on the 3′ side (*see* Fig. 1b). In order to have an efficient self-circularization of DNA molecules during the second ligation step, the digestion of your viewpoint by these two enzymes must give rise to a DNA fragment of at least 300 bp.

One of the two restriction enzymes will be used during the first digestion step of the 4C experiment and is called the first cutter. It is this site that will capture the interactions occurring at the viewpoint. In order to be used as first cutter, a restriction enzyme needs to be able to digest chromatinized DNA and its activity should not be blocked by methylated CpG. Typical first cutters meeting these criteria are DpnII, Csp6I, and NlaIII.

The other restriction enzyme will be used during the second digestion step and is designed as second cutter. This cutter will help reduce the size of the captured region, which will decrease amplification biases. As for the first cutter, it should not be sensitive to methylated CpG. However, because it will be used to digest purified DNA there is no requirement regarding its ability to digest chromatinized DNA. The following restriction enzymes can be used as second cutters: DpnII, Csp6I, NlaIII, and BfaI.

It is possible to select other first and second cutters but make sure that they generate cohesive ends to have efficient ligations. Moreover, their restriction sites should not be composed of repeats (e.g., CCTT), which tend to introduce biases because of the poor distribution of those sites within the genome. In this protocol we use DpnII as first cutter and BfaI as second cutter. If different cutter sets are to be used, the protocol is the same except that the restriction enzyme inactivation conditions might have to be adjusted accordingly.

3.2 Designing Primers Specific to Your Viewpoint

For the PCR step of the 4C experiment, a set of primers specific to your viewpoint need to be designed (*see* Fig. 1b). The viewpoint-specific forward primer should be between 18 and 20 nucleotides long and its 3′ end has to overlap and finish on the first cutter site. The viewpoint-specific reverse primer should be 18–20 nucleotides long. It can be anywhere within 100 bp of the second cutter restriction site but should be as close as possible to this site. As for the design of any PCR primers, some general considerations have to been taken into account. The primers should have a melting temperature around 55 °C; if a different melting temperature is used the PCR program must be modified accordingly. The difference between the melting temperature of both primers should be within a ±3 °C range. The percentage of GC must be comprised between 30 and 70 %. Primers with base repeats and primers located within repeated regions of the genome are proscribed.

In order to be compatible with the Illumina deep sequencing TruSeq chemistry, Illumina adaptors are added to the viewpoint-specific forward and reverse primers (*see* Fig. 1b). The Illumina sequencing reaction will start from the sequencing adaptor located on the forward primer, which is referred to as the reading primer. It is important to note that the viewpoint-specific forward primer will always be the first sequence to be read during the sequencing reaction. This means that this sequence can be used to sort reads obtained from multiplexed 4C-seq experiments performed at different viewpoints. However, in the case of multiplexed 4C-seq experiments performed at the same viewpoint, a tag can be added between the sequencing adaptor and the viewpoint-specific forward primer to be able to differentiate reads (*see* Fig. 1b).

When your viewpoint-specific primers are designed, assemble and order the following 4C primers:

4C Reading Primer

5′-AATGATACGGCGACCACCGA-ACACTCTTTCC CTACACGACGCTCTTCCGATCT-(optionaltag)-(viewpoint-specific forward primer sequence)-3′

4C Reverse Primer

5′-CAAGCAGAAGACGGCATACGA-(viewpoint-specific reverse primer sequence)-3′

This reverse primer is for single-end sequencing; *see* **Note 1** for the sequence of a TruSeq Paired-End compatible reverse primer.

3.3 Cell Fixation

1. Grow ten million cells. Before fixation, make sure that the cells are in the exponential growing phase.
2. For cells in suspension transfer the cells into a 15 ml tube, spin for 4 min at 600 × *g*, remove the supernatant, and resuspend the cells in 9.5 ml of medium.

   For adherent cells wash the cells two times with PBS, add 2 ml of trypsin, incubate for 5 min at 37 °C, then add 7.5 ml of medium, and transfer to a 15 ml tube (*see* **Note 2)**.
3. Add fresh formaldehyde to a final concentration of 2 %, e.g., 550 µl of 37 % formaldehyde. Mix immediately and incubate for 10 min at room temperature on a rotator (*see* **Note 3**).
4. Stop the cross-linking by adding glycine to a final concentration of 125 mM, e.g., 525 µl of 2.5 M glycine.
5. Immediately centrifuge for 8 min at 600 × *g* at 4 °C, then remove all the supernatant, and place the tube on ice.

3.4 First Cutter Digestion

1. Resuspend the pellet in 1 ml of freshly prepared ice-cold lysis buffer, transfer to a 1.5 ml tube, and incubate for 10 min on ice.
2. To check the efficiency of the lysis reaction, take a 3 µl aliquot of the cell suspension and combine with 3 µl of Methyl

Green-Pyronin staining. Mount the mixture on a microscope slide. If the lysis was efficient the nucleus should appear blue-green with very little red-pink (RNA). If the lysis was not complete, and the cytoplasm is still present, cells will appear red-pink with blue-green nucleus. Efficient lysis is critical for ensuring an efficient first digestion. If the lysis is not complete you can increase the incubation time or use a douncer.

3. Centrifuge the cell suspension for 5 min at 750 × *g* at 4 °C and remove the supernatant.
4. Resuspend the nuclei in 510 µl of 1× DpnII restriction buffer.
5. Incubate at 37 °C for 15 min.
6. Add 7.5 µl of 20 % SDS and incubate for 1 h at 37 °C while shaking at 900 rpm using an Eppendorf Thermomixer (*see* **Note 4**).
7. Do some up-and-down pipetting to disrupt cell aggregates that might have formed during the incubation with SDS.
8. Add 60 µl of 25 % Triton X-100 and incubate for 1 h at 37 °C while shaking at 900 rpm (*see* **Note 5**).
9. Do some up-and-down pipetting to disrupt cell aggregates that might have formed during the incubation with Triton X-100.
10. Collect a 5 µl aliquot and store it at 4 °C. It will be used as a negative control for the digestion.
11. Add 400 U of DpnII and incubate overnight at 37 °C while shaking at 900 rpm.
12. Add 200 U of DpnII and incubate for 4 h at 37 °C while shaking at 900 rpm.
13. To assess the digestion efficiency:
    (a) Collect a 5 µl aliquot of digested sample. For the following steps (**steps 13a–13e**), also process the first digestion negative control collected at **step 10**.
    (b) Add 95 µl of 10 mM Tris–HCl pH 7.5 to the aliquots.
    (c) Add 5 µl of RNase A (10 mg/ml) and incubate for 30 min at 55 °C.
    (d) Add 5 µl of Proteinase K (10 mg/ml) and incubate for 1 h at 65 °C.
    (e) Run 20 µl of non-digested and digested control samples on a 0.6 % agarose gel. If the digestion is complete the digested sample should appear as a smear ranging from 800 to 5000 bp with the maximum amount of DNA around 2000 bp (*see* Fig. 2a).
14. If the first digestion is incomplete repeat **steps 12** and **13**. If the digestion is complete, heat-inactivate DpnII by incubating

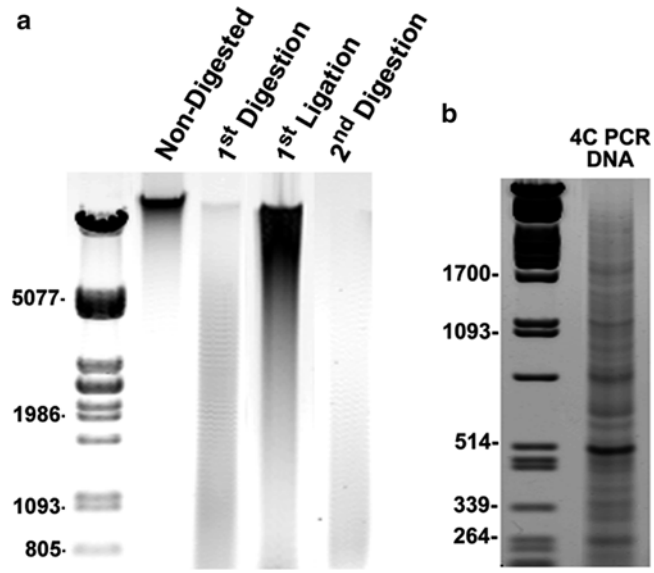


Fig. 2 (**a**) 4C-seq experiment digestion and ligation controls run on a 0.6 % agarose gel. (**b**) PCR product of a 4C-seq experiment ran on a 1.5 % agarose gel. The most intense band corresponds to the non-cut PCR product

the sample for 30 min at 65 °C under agitation at 900 rpm (if you used a different first cutter that cannot be heat inactivated *see* **Note 6**).

3.5 First Ligation

1. Transfer the digested sample to a 50 ml tube.
2. Add to the sample: 5.7 ml of H_2O, 700 µl of 10× ligase buffer, 50 U of T4 DNA ligase.
3. Mix gently and incubate overnight at 16 °C.
4. Determination of the ligation efficiency:
   (a) Collect a 100 µl aliquot of ligated sample.
   (b) Add 5 µl of RNase A (10 mg/ml) and incubate for 30 min at 55 °C.
   (c) Add 5 µl of Proteinase K (10 mg/ml) and incubate for 1 h at 65 °C.
   (d) Collect the aqueous phase and run 20 µl on a 0.6 % agarose gel along with the first digestion sample from Subheading 3.4, **step 13e**.
5. If the ligation was efficient, the ligated sample should appear as a large band of high molecular mass; proceed to **step 6** (*see* Fig. 2a). If this is not the case, add ATP to a final concentration of 1 mM and 25 U of T4 DNA ligase, and repeat **steps 3** and **4**.
6. To reverse the cross-link, add 30 µl Proteinase K (10 mg/ml) and incubate overnight at 65 °C.

7. Add 30 µl of RNase A (10 mg/ml) and incubate for 1 h at 55 °C. If nucleus aggregates are still present after this incubation step, add 30 µl of Proteinase K (10 mg/ml) and incubate for 2 h at 55 °C.
8. Add 7 ml of phenol-chloroform, mix vigorously, and centrifuge for 15 min at 3000 × *g* at room temperature.
9. Transfer the aqueous phase to a new 50 ml tube and add: 8 ml of H_2O, 1 ml 3 M NaAc pH 5.2, 7 µl glycogen (20 mg/ml), and 35 ml of 100 % ethanol.
10. Mix thoroughly and incubate the sample at −80 °C until frozen.
11. Centrifuge for 45 min at 9500 × *g* at 4 °C.
12. Remove the supernatant and wash the pellet with 10 ml of ice-cold 70 % ethanol.
13. Centrifuge for 15 min at 3000 × *g* at 4 °C.
14. Discard the supernatant and dry the pellet.
15. Dissolve the pellet in 150 µl of 10 mM Tris–HCl pH 7.5 at 37 °C. After this step, the samples can be stored at −20 °C.

3.6 Second Cutter Digestion

1. To the 150 µl of sample from Subheading 3.5, **step 15**, add 300 µl of H_2O, 50 µl 10× of NEB CutSmart, and 50 U of BfaI.
2. Incubate overnight at 37 °C.
3. Determine the second digestion efficiency:
   (a) Collect a 10 µl aliquot of sample digested with the second cutter.
   (b) Add 90 µl 10 mM Tris–HCl pH 7.5.
   (c) Run 20 µl on a 0.6 % agarose gel along with the first ligation control sample from Subheading 3.5, **step 4d**. If the second digestion is complete the sample should appear as a smear with the maximum intensity around 1000 bp (*see* Fig. 2a).
4. If the digestion is incomplete add 25 U of BfaI, incubate 4 h at 37 °C and repeat **step 3**. If the digestion is complete, heat-inactivate BfaI by incubating the sample for 30 min at 80 °C under agitation at 900 rpm (if you use a different second cutter that cannot be heat inactivated *see* **Note 7**).

3.7 Second Ligation

1. Transfer the sample to a 50 ml tube and add 12.1 ml of H_2O, 1.4 ml 10× ligation buffer, and 100 U of T4 DNA ligase.
2. Incubate overnight at 16 °C.
3. To precipitate the DNA add 460 µl 3 M NaAc pH 5.2, 7 µl glycogen (1 mg/ml), and 35 ml of 100 % ethanol.
4. Mix thoroughly and incubate the sample at −80 °C until frozen.

5. Centrifuge for 45 min at 9500 × *g* at 4 °C.
6. Remove the supernatant and wash the pellet with 10 ml of ice-cold 70 % ethanol.
7. Centrifuge for 15 min at 3000 × *g* at 4 °C.
8. Discard the supernatant and dry the pellet.
9. Dissolve the pellet in 150 µl 10 mM Tris–HCl pH 7.5 at 37 °C.
10. Purify the samples using the QIAquick PCR purification kit. Follow the kit instructions but use three columns per sample. Elute each column with 50 µl of EB buffer and pool identical samples.
11. Measure the DNA concentration in the pooled samples using a Nanodrop spectrophotometer. Typically, the total amount of DNA lies between 40 and 60 µg for ten million cells.
12. The 4C DNA can be stored at –20 °C or directly used for PCR.

3.8 PCR

1. Prepare aliquots of the 4C primers at 20 µM in 10 mM Tris–HCl pH 7.5.
2. Determine the linear range of amplification of the primer sets by performing PCRs using different amounts of 4C DNA: 25, 50, 100, 200, and 400 ng.
   (a) Assemble the following reaction mix to perform the test PCRs:
   - 10× ELT PCR buffer 1: 2.5 µl.
   - 10 mM dNTP: 0.5 µl.
   - 20 µM Reading primer: 3 µl.
   - 20 µM Reverse primer: 3 µl.
   - ELT Pol (5 U/µl): 0.35 µl.
   - 4C DNA template: 25–400 ng.
   - H_2O: complete to 25 µl.

   (b) Perform the PCR using the following program: 3 min at 95 °C; (10 s at 95 °C; 1 min at 55 °C; 3 min at 68 °C) × 29 cycles; 5 min at 68 °C; ∞ at 4 °C.
   (c) Run 15 µl of the PCR products on a 1.5 % agarose gel and quantify to assess if the amplification was linear. Typically, 4C PCR products form a smear composed of several faint bands and a strong one corresponding to the non-digested first cutter site (*see* Fig. 2b).
3. If the test PCRs yield satisfying results, prepare the following PCR master mix:
   - 10× ELT PCR buffer 1: 80 µl
   - 10 mM dNTP: 16 µl

- 20 μM Reading primer: 54 μl
- 20 μM Reverse primer: 54 μl
- 4C DNA template: 3.2 μg
- ELT Pol (5 U/μl): 11.2 μl
- H_2O: Complete to 800 μl

4. Split the PCR master mix into 16 reactions of 50 μl each and perform the PCRs using the same program as in **step 2b**.
5. Pool all identical PCR reactions in a 15 ml tube and purify the 4C PCR products using the QIAquick PCR Purification Kit from Qiagen. Follow the kit instructions but use three columns per sample. Elute the 4C PCR products with 50 μl of elution buffer and pool identical samples (for a total volume of 150 μl per sample).
6. To remove primers and primer dimers, purify the sample using Agencourt AMPure XP. Make sure that your sample volume is exactly 150 μl; if it is not the case adjust the volume to 150 μl with H_2O. Add 135 μl of Agencourt AMPure XP to the sample (0.9 volume ratio) and follow the manufacturer's protocol. Elute the 4C PCR DNA using 150 μl of 10 mM Tris–HCl pH 7.5.
7. Quantify the sample using a Nanodrop. Typically you should obtain from 10 to 20 μg of 4C PCR product. Because sample purity is important for the deep sequencing step, check that the absorbance ratio A260/A280 is close to 1.80 and that the A260/A230 ratio is above 1.5.
8. Assess the quality of the purified 4C PCR products by running 300 ng on a 1.5 % agarose gel.
9. The samples are ready to be deep sequenced and can be stored at −20 °C.

3.9 Deep Sequencing

Using the primers described in this protocol the samples can be sequenced on any Illumina sequencing platform compatible with Single-End TruSeq chemistry. Because the adaptors for deep sequencing are incorporated to the 4C reading and reverse primers, no further manipulation of the 4C PCR product is needed prior to deep sequencing.

Typically, 1.5 millions reads are required per 4C sample; it is thus technically possible to multiplex more than 100 4C samples on one lane of a HiSeq 2000 (200 million reads per lane). However, an important point to take into account for the multiplexing of 4C samples is the clustering process during deep sequencing. This step determines where the DNA molecules are located on the sequencer flow cell. To be optimal, the clustering requires good sequence diversity for the first bases to be sequenced. When sequencing 4C PCR products, make sure to multiplex several experiments and verify that the viewpoint-specific forward primers (and tags if

added) do not share similar nucleotides at similar positions. Alternatively, it is possible to avoid this problem by multiplexing 4C experiments with other types of deep sequencing experiments (e.g., ChIP-seq or RNA-seq), or with the PhiX control library.

Because the tag and the 4C reading primer are sequenced before the unknown interacting sequence, we recommend sequencing at least 50 nucleotides. This allows to sequence a minimum of 29 nucleotides of the unknown interacting sequence (50 bp–3 bp tag–18 bp 4C primer = 29 bp).

3.10 4C-Seq Data Analysis

The analysis of 4C-seq data is generally carried out as follows (*see* Fig. 3):

1. De-multiplexing the 4C reads: This step consists of separating multiplexed samples and removing tag and primer sequences from the reads. For the sorting, reads are binned using the different viewpoint-specific primer sequences and tags. Usually, for this process no mismatch is allowed for the tag while one mismatch can be tolerated for the primer sequences (as long as the primer sequences are sufficiently different). The tags and the primer sequences are removed but the first cutter site sequence is kept. At the end of this process, each bin contains the reads obtained for a specific viewpoint and a specific tag, and each read should start with the sequence of the first cutter restriction site.
2. Mapping the 4C reads: Prior to mapping, reads that correspond to the uncut and self-ligated fragments are removed. These sequences can represent an important portion of the reads and need to be removed in order to avoid biases during the analysis. Subsequently, read sequences that do not originate from ligation events between two first cutter sites are also removed. To this aim, reads are mapped to an in silico library comprising all the unique restriction fragments generated by the combination of the first and second cutter. Only reads that match those fragments are mapped. A wiggle track can then be generated and visualized using the UCSC genome browser (https://genome.ucsc.edu/) or the Integrative Genomics Viewer (IGV) software (http://www.broadinstitute.org/igv/).
3. 4C profiles and statistics: An important notion to bear in mind when interpreting 4C experiments is that an interaction between a region of interest and a specific locus will be captured over several first cutter restriction sites. Consequently, a single restriction site with a very high signal does not likely reflect a real interaction but rather an artifact. To avoid these biases, a running mean or median is used to smooth the data and decrease the significance of outlier data points. A window of 11 restriction sites or 3 kbp gives good results when a four cutter is used as first cutter. More elaborated analyses can also

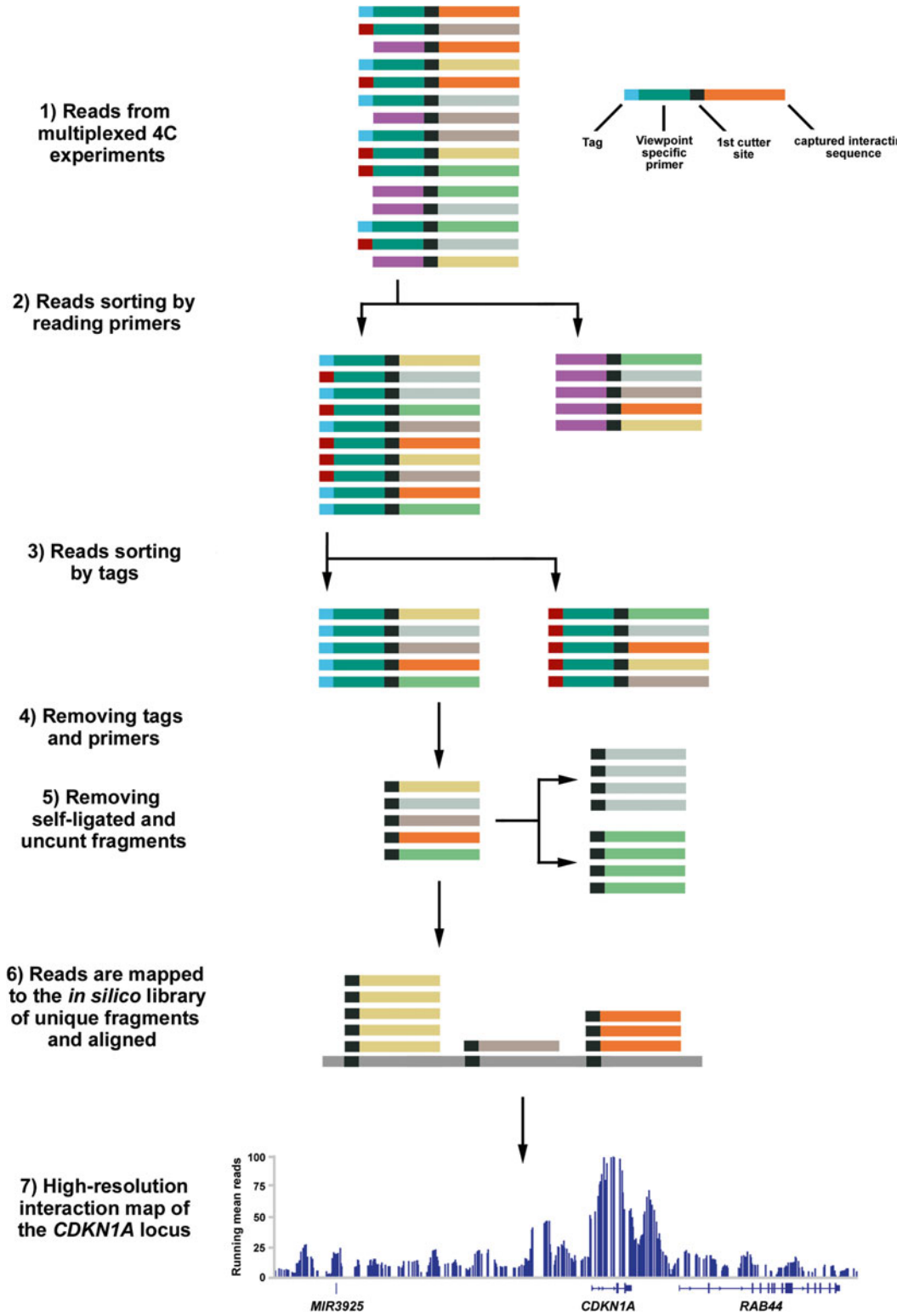


Fig. 3 General analysis workflow of multiplexed 4C-seq experiments. The workflow is shown in full for the sample with the *green reading primer* and *blue tag*. The high-resolution map of the *CDKN1A* locus spans 200 kbp

be performed such as the calculation of *z*-scores, which allows one to use the false discovery rate and identify nonrandom 4C-signal [9]. Finally, to correct for the decrease in signal intensity as the distance from the viewpoint increases, 4C domainogram algorithms have also been developed [9, 10].

Two pipelines are available to analyze 4C-seq data as described in the previous paragraph. The first was developed conjointly by Amos Tanay's research group and Wouter De Laat's research group (http://compgenomics.weizmann.ac.il/tanay/?page_id=367) [11]. The second pipeline was developed by the École polytechnique fédérale de Lausanne's Bioinformatics and Biostatistics Core Facility Python Library team (http://bbcf.epfl.ch/bbcflib/index.html) [12]. Both have been made available to the research community free of charge.

4 Notes

1. 4C reverse primer for paired-end deep sequencing: 5′-CAA GCA GAA GAC GGC ATA CGA GAT NNNNNN G TGA CTG GAG TTC AGA CGT GTG CTC TTC CGA TC-(region of interest reverse primer sequence)-3′.
   NNNNNN = TruSeq barcode sequence
2. It is recommended to perform the cross-linking step with cells in suspension as they yield better results.
3. Formaldehyde concentration during cross-linking might need to be adjusted for best results and typically ranges between 1 and 3 %.
4. The SDS helps to remove uncross-linked protein from the DNA to increase digestion efficiency.
5. The Triton X-100 quenches the SDS prior to the first digestion.
6. If you cannot heat inactivate your first cutter, proceed as follows:
   (a) Add 40 μl of 20 % SDS and incubate for 30 min at 65 °C.
   (b) Transfer the sample to a 50 ml tube and add: 5.4 ml of H_2O, 700 μl 10× ligase buffer, and 375 μl 20 % Triton X-100.
   (c) Incubate for 1 h at 37 °C.
   (d) Add 50 U of T4 DNA ligase and resume the experiment as of Subheading 3.5, **step 3**.
7. If you cannot heat inactivate your second cutter, proceed as follows:
   (a) Add 500 μl phenol-chloroform and mix vigorously.
   (b) Centrifuge at 16,000 × g for 10 min.

(c) Transfer the aqueous phase to a new 1.5 ml tube.

(d) Add: 33 µl of 3 M NaAc pH 5.6 and 970 µl of 100 % ethanol.

(e) Incubate at −80 °C for 30 min.

(f) Centrifuge at 16,000 × g for 30 min at 4 °C.

(g) Wash the DNA pellet with 150 µl of cold 70 % ethanol.

(h) Centrifuge at 16,000 × g for 5 min at 4 °C.

(i) Resuspend the pellet in 500 µl 10 mM Tris–HCl pH 7.5 and resume experiment as of Subheading 3.7, **step 1**.

Acknowledgements

We are very grateful to Dr. Patrick Wijchers and Dr. Wouter de Laat for giving us the opportunity to learn the 4C-seq technique. We also are very thankful to Dr. Amos Tanay and to the EPFL's Bioinformatics and Biostatistics Core Facility Python Library team for having made the 4C-seq pipelines available to the scientific community. We also want to thank Dr. Benoit Guillemette and Jennifer Chambers for their constructive comments on earlier drafts of this chapter. This work was supported by the Canada Research Chair Program "Mechanisms of Gene Transcription" to LG.

References

1. Merkenschlager M, Odom DT (2013) CTCF and cohesin: linking gene regulatory elements with their targets. Cell 152(6):1285–1297
2. de Laat W, Duboule D (2013) Topology of mammalian developmental enhancers and their regulatory landscapes. Nature 502(7472): 499–506
3. Bulger M, Groudine M (2010) Enhancers: the abundance and function of regulatory sequences beyond promoters. Dev Biol 339(2): 250–257
4. Kagey MH, Newman JJ, Bilodeau S et al (2010) Mediator and cohesin connect gene expression and chromatin architecture. Nature 467(7314):430–435
5. Montavon T, Soshnikova N, Mascrez B et al (2011) A regulatory archipelago controls Hox genes transcription in digits. Cell 147(5):1132–1145
6. Dekker J, Rippe K, Dekker M et al (2002) Capturing chromosome conformation. Science 295(5558):1306–1311
7. Simonis M, Klous P, Splinter E et al (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38(11):1348–1354
8. Zhao Z, Tavoosidana G, Sjolinder M et al (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 38(11):1341–1347
9. Splinter E, de Wit E, van de Werken HJ et al (2012) Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. Methods 58(3):221–230
10. de Wit E, Braunschweig U, Greil F et al (2008) Global chromatin domain organization of the Drosophila genome. PLoS Genet 4(3):e1000045
11. van de Werken HJ, Landan G, Holwerda SJ et al (2012) Robust 4C-seq data analysis to screen for regulatory DNA interactions. Nat Methods 9(10):969–972
12. Noordermeer D, Leleu M, Splinter E et al (2011) The dynamic architecture of Hox gene clusters. Science 334(6053):222–225

Chapter 17

Global Mapping of Open Chromatin Regulatory Elements by Formaldehyde-Assisted Isolation of Regulatory Elements Followed by Sequencing (FAIRE-seq)

Stéphanie Bianco, Sébastien Rodrigue, Bruce D. Murphy, and Nicolas Gévry

Abstract

Genetic information is organized in a complex structure composed of DNA and proteins together designated chromatin. Chromatin plays a dynamic role in transcriptional processes in that alteration of the interaction between its components results in the deregulation of cellular transcriptional program. Modification of epigenetic marks, variation in the precise positioning of nucleosomes, and consequent mobilization of nucleosomes regulate the access of various transcriptional factors to its underlying DNA template. Nucleosome-depleted regions, also designated open chromatin domains, are associated with active DNA regulatory elements, including promoters, enhancers, silencers, and insulators. Here, we describe the protocol of a rapid and simple technique entitled FAIRE (formaldehyde-assisted isolation of regulatory elements). Combined with high-throughput sequencing (FAIRE-seq), this procedure allows isolation of nucleosome-free regions and their mapping along the genome, thereby providing a global view of cell-specific regulatory elements.

Key words Chromatin, Formaldehyde, Regulatory elements, Next-generation sequencing, Protein–DNA interaction, Epigenetics, Chromatin accessibility, Nucleosome-depleted regions

1 Introduction

Gene expression is a highly regulated process that may be modulated at several levels, resulting in the conversion of genetic information to functional protein synthesis. Transcription is the first step of this complex process, employing RNA polymerase II, the main component of the basal transcription machinery, acting at the core promoter. It works in combination with a vast number of transcription factors that bind distal regulatory elements and mediate enhancer-promoter communication. Transcription factor binding is facilitated by eviction or destabilization of nucleosomes from chromatin. Nucleosome-depleted regions, also known as open chromatin domains, are associated with active

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_17,

DNA regulatory elements, including promoters, enhancers, silencers, and insulators.

Nucleosome stability is controlled by at least three different mechanisms: the action of chromatin remodeling complexes that contain an ATPase subunit [1]; the posttranslational modification of histones such as acetylation, methylation, phosphorylation, sumoylation, and ubiquitination [2]; and the replacement of canonical histones with histone variants [3]. As a consequence, both the interaction between DNA and nucleosome and the accessibility of DNA to transcription factors are highly dynamic.

Initially, nucleosome-depleted regions were identified by their hypersensitivity to nuclease cleavage, either DNase I or MNase, that were subsequently mapped along the genome by recently developed high-throughput sequencing technology and computational analyses [4, 5]. DNA regulatory elements can also be isolated by chromatin immunoprecipitation (ChIP) employing antibodies against a specific transcription factor and processed by massively parallel sequencing (ChIP-seq) to determine protein interaction with DNA along the genome. ChIP-seq procedure necessitates having antibody with capability to attach to a specific transcription factor, thereby allowing for precise mapping of the position of this transcription factor. However ChIP-seq does not provide a global view of the complexities of the dynamic control of chromatin function.

The purpose of this chapter is to describe a simple and rapid technique, named FAIRE (formaldehyde-assisted isolation of regulatory elements) which allows the isolation and the mapping of nucleosome-free regions accessible by transcription factors. This is an alternative method that does not require enzyme or antibody.

This method has been successfully used to define open chromatin regions in normal and pathologic cell models, to provide clues about transcription factors associated with nucleosome-free regions during cell differentiation and to identify sequence variation (SNP) in human disease [6–9].

Briefly, DNA-protein complexes are cross-linked with formaldehyde. Then, cells are lysed and the chromatin is fragmented by sonication to obtain fragments less than 500 bp long. Sheared chromatin is extracted by phenol/chloroform to separate nucleosome-free region from total chromatin. DNA is then purified, and the quality tested by quantitative PCR (qPCR) analysis using candidate regions of known abundant and moderate enrichment, and quantified by spectrophotometry or fluorometric assay before library preparation. Following blunt-ended DNA repair, a size selection is performed to recover average fragments of 300 bp. This step is achieved by the solid-phase reversible immobilization (SPRI) beads. SPRI beads are carboxyl coated magnetic particles that have the advantage to bind double stranded and single stranded DNA in a rigorously size-dependent manner according to

the volume ratio of SPRI bead and DNA solution [10]. This procedure has the advantage compared to classical agarose gel DNA band excision in that it avoids cross-contamination, each library being prepared in independent tube. Furthermore, SPRI beads replace traditional DNA cleanup methods by allowing elimination of excess of salts, enzymes, nucleotides, primers, primer dimers, and self-ligated product in a single washing step. Following size selection, DNA is subjected to adapter ligation and amplified using barcoded primers. In overview, DNA samples are tested for purity, quality, and size prior to be mixed for highly multiplexed DNA sequencing (*see* Fig. 1).

2 Materials

2.1 FAIRE DNA Isolation

1. 1.1 % PBS-formaldehyde mix from 37 % formaldehyde and PBS 1× of pH 7.4.
2. Glycine 2.5 M.
3. SDS lysis buffer: 1 % SDS, 10 mM EDTA, 50 mM Tris–HCI of pH 8.1. Add 1 mM PMSF and 1× protease inhibitor cocktail before use.
4. 100× protease inhibitor cocktail stock solution in 90 % ethanol: 0.2 mM pepstatin A, 72 μM leupeptin, and 26 μM aprotinin.
5. Bioruptor sonication system (Diagenode).
6. Phase lock gel (PLG heavy, Eppendorf).
7. Phenol/chloroform/isoamyl alcohol (25:24:1).
8. QIAquick PCR or MinElute Purification Kit (Qiagen).
9. Nanodrop or Quant-iT Picogreen dsDNA Assay Kit (Invitrogen, P11496).
10. DNase-free RNaseA (10 μg/μL).
11. Proteinase K (20 μg/μL).

2.2 Library Preparation Components

1. SPRI beads, AMPure XP, Agencourt (Beckman Coulter, A63881).
2. Magnetic support for 0.2 mL PCR tube.
3. 0.2 mL PCR tube.
4. Filter tips.
5. Molecular grade sterile water.
6. Freshly prepared 70 % EtOH.
7. Annealed adapter A-B (*see* Table 1).
8. Illumina barcoded primer (*see* Table 1).
9. End-Repair Mix and 10× buffer (Enzymatics, Y914-HC-L).

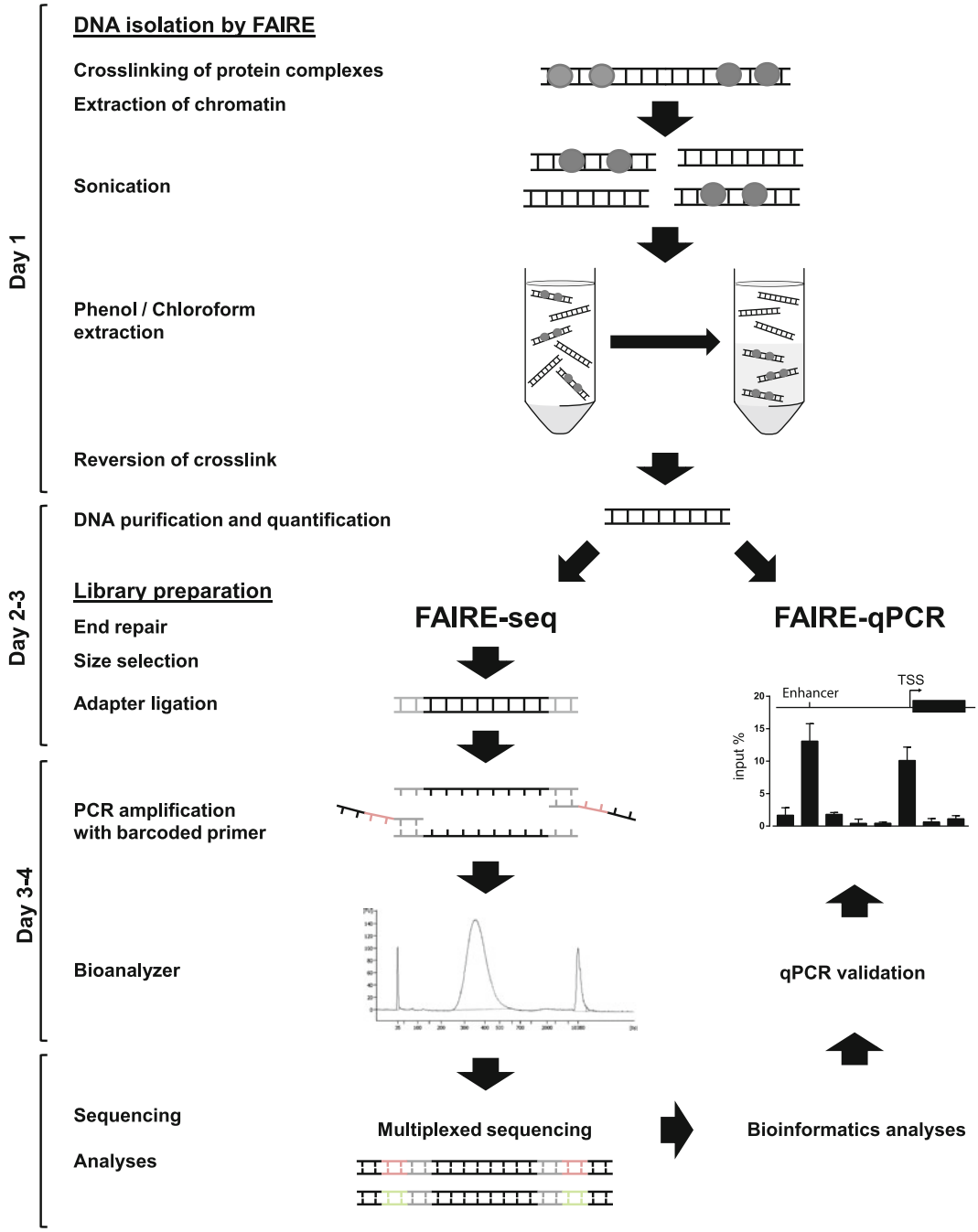


Fig. 1 FAIRE-seq procedure. A step-by-step protocol of FAIRE-seq showing the major procedures and typical timeline for the carrying out the experiment

10. dNTPs.
11. T4 DNA ligase and 2× buffer (Enzymatics, L603-HC-L).
12. Taq-B DNA polymerase and 10× buffer (Enzymatics, P725L).

Table 1
Oligonucleotide sequences for library preparation

Name	Sequence 5′–3′
Adapter A up	/5AmMC6/ACACTCTTTCCCTACACGACGCTCTTCCGATCT
Adapter A down	AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGTAC/3AmMO/
Adapter B up	AGATCGGAAGAGCACACGTCTGAACTCCAGTC/3AmMO/
Adapter B down	/5AmMC6/GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT
Forward primer	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTC TTCCGATCT
Reverse primer	CAAGCAGAAGACGGCATACGAGAT-**NNNNNN**-GTGACTGGAGTTCAGA CGTGTGCTCTTCCGATC

5AmMC6: 5′ amino modifier C6
3AmMO: 3′ amino modifier
NNNNNN: barcode sequence
PCR primers are compatible with Illumina technology

13. Phusion High-Fidelity PCR Master Mix with HF Buffer (NEB, M0531S).
14. SYBR Green I 10×.
15. Thermal cycler.
16. Real-time quantitative PCR system.
17. Agilent 2100 bioanalyzer and the High Sensitivity DNA kit.

3 Methods

This protocol is described for adherent human cells cultured in 150 mm dishes at 70 % confluence. It has also been successfully tested with frozen tissues. For different use, some steps would necessitate adaptations.

3.1 Cross-Link of DNA–Protein Complexes with Formaldehyde

1. Remove culture medium from cells.
2. Add 20 mL PBS-formaldehyde mix (1.1 %) per 150 mm dish and incubate for 10 min at room temperature.
3. Add 1 mL glycine 2.5 M (final concentration 125 mM) and incubate for 5 min at room temperature.
4. Remove formaldehyde and wash cells twice with cold PBS.
5. Harvest cells by scrapping into 500 μL of ice cold PBS and transfer to 1.5 mL tubes.
6. Centrifuge for 6 min at 850 × *g* and remove supernatant.
7. Cells can be snap frozen in liquid nitrogen at this point and stored at −80 °C.

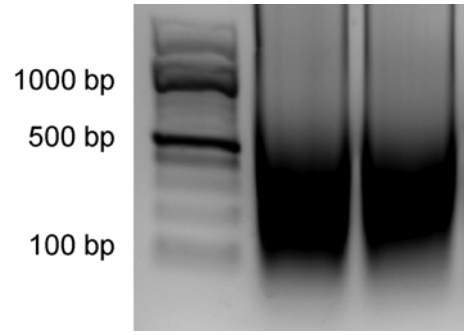


Fig. 2 Migration patterns of sonicated DNA on 1.2 % agarose gel electrophoresis. The ideal size of DNA fragments following sonication is between 200 and 500 bp

3.2 Cellular Lysis and Chromatin Preparation

1. Resuspend cells in 300 µL of freshly prepared SDS lysis buffer and incubate for 30 min on ice.
2. Sonicate samples in Bioruptor on "high"-intensity position for 30 s ON and 30 s OFF for 15 cycles at 4 °C (*see* **Notes 1** and **2**).
3. Centrifuge for 10 min at full speed (18,000 × *g*) at 4 °C to clear cellular debris and transfer supernatant to a new tube.
4. Take a small aliquot (5 % of the volume) to verify fragment size on agarose gel (*see* **Note 2**, Fig. 2).
5. Keep 5 % of the volume as input for analysis by qPCR.

3.3 Phenol/ Chloroform Extraction of Nucleosome-Depleted Regions

1. Add an equal volume of phenol:chloroform to sample into a phase-lock gel (*see* **Note 3**).
2. Vortex and spin for 1 min at 18,000 × *g* at room temperature. Transfer the upper phase to new tubes.
3. Perform **steps 1** and **2** two other times.

3.4 DNA Purification and Quantification

1. Reverse formaldehyde cross-linking by incubating samples overnight at 65 °C (*see* **Note 4**).
2. Incubate samples with 20 µg of RNase A for 30 min at 37 °C and then add 100 µg of proteinase K for 1 h at 65 °C.
3. Purify DNA using QIAquick PCR or MinElute purification kit (*see* **Note 5**).
4. DNA obtained can be analyzed using qPCR (*see* **Note 6**).
5. Quantify DNA with Nanodrop or Quant-iT Picogreen dsDNA Assay Kit.

3.5 DNA End Repair for Library Preparation

1. Mix 50 ng of DNA with end repair mix in a total volume of 25 µL in 0.2 mL PCR tube:

   1–19 µL ADN (50 ng).

   2.5 µL 10× End repair buffer.

2.5 μL 1 mM dNTP mix.

1 μL End repair enzyme mix.

H_2O (total 25 μL).

2. Incubate for 30 min at room temperature.
3. Purify DNA with solid-phase reversible immobilization (SPRI) magnetic beads (AMPure XP) (*see* **Note 7**).
4. Add 25 μL of beads and mix tenfold by pipetting (*see* **Note 8**).
5. Incubate for 5 min at room temperature.
6. Place the tube 2–3 min on the magnetic support.
7. Remove supernatant and keep the tube on magnetic support.
8. Add 100 μL of 70 % EtOH, wait for 30 s, and remove EtOH (*see* **Note 9**).
9. Repeat **step 8** once.
10. Incubate the beads at room temperature for 10–15 min to dry residual EtOH (*see* **Note 10**).
11. Remove the tube from magnetic support.
12. Resuspend the beads with 52 μL of water.
13. Incubate at room temperature for at least 1 min.
14. Place the tube 2 min on the magnetic support and transfer supernatant (50 μL) to a new tube.

3.6 Size Selection

1. Add 35 μL of beads (bead/DNA ratio of 0.7) and mix tenfold by pipetting.
2. Incubate for 5 min at room temperature (*see* **Note 11**).
3. Place the tube 2–3 min on the magnetic support.
4. Transfer supernatant (~82 μL) to a new tube.
5. Add 12.3 μL of beads (bead/DNA ratio of 0.15) and mix tenfold by pipetting (*see* **Note 12**).
6. Perform **steps 5–11** from Subheading 3.5.
7. Resuspend the beads with 14 μL of water.
8. Incubate at room temperature for at least 1 min.
9. Place the tube 2 min on the magnetic support and transfer supernatant (12 μL) to a new tube.

3.7 Adapter Ligation

1. Mix DNA with ligation reaction by pipetting.

   12 μL DNA.

   15 μL 2× ligase buffer.

   1 μL Annealed adapter A-B (*see* **Note 13**).

   2 μL T4 DNA ligase.
2. Incubate for 10 min at room temperature.

3. Add 20 µL of water to decrease polyethylene glycol (PEG) concentration from ligase buffer.
4. Purify DNA with 50 µL of SPRI magnetic beads (bead/DNA ratio of 1) performing **steps 5–11** from Subheading 3.5 (*see* **Note 5**).
5. Resuspend the beads with 23.5 µL of water.
6. Incubate at room temperature for at least 1 min.
7. Place the tube 2 min on the magnetic support and transfer supernatant (21.5 µL) to a new tube.

3.8 Nick Translation

1. Mix DNA with nick translation reaction by pipetting.

   21.5 µL ADN.

   2.5 µL 10× buffer.

   0.5 µL dNTPs 10 mM.

   0.5 µL Taq B DNA pol.
2. Incubate for 20 min at 66 °C.
3. Purify DNA with 25 µL of SPRI magnetic beads (bead/DNA ratio of 1.0) performing **steps 5–11** from Subheading 3.5 (*see* **Note 5**).
4. Resuspend the beads with 15 µL of water.
5. Incubate at room temperature for at least 1 min.
6. Place the tube 2 min on the magnetic support and transfer supernatant (13 µL) to a new tube.

3.9 Amplification

1. Split DNA for three PCR reactions and mix with the following reagents:

   4.3 µL DNA.

   12.5 µL Master mix Phusion HF.

   0.5 µL SYBR Green 10×.

   1 µL Forward primer 10 µM (*see* Table 1).

   1 µL Reverse primer with barcode 10 µM (*see* Table 1).

   5.7 µL H_2O (total 25 µL).
2. Amplify the samples using the following program:

Denature at 98 °C for 30 s	
98 °C—10 s 65 °C—30 s 72 °C—30 s	×(*n*) cycles.

3. Stop reaction during the exponential phase (*see* **Note 14**).

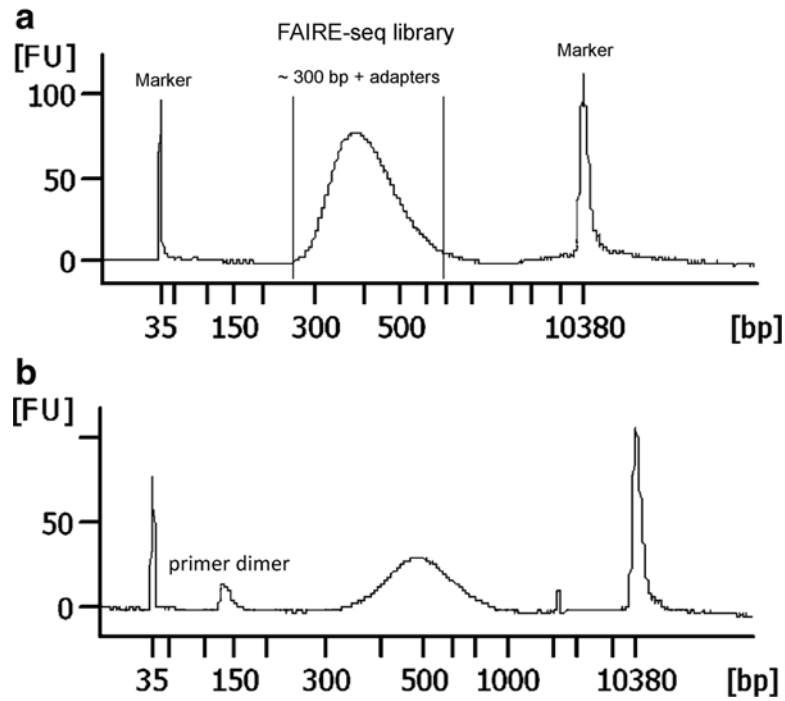


Fig. 3 Bioanalyzer analyses: (**a**) The ideal DNA fragment distribution following FAIRE-seq library preparation is around 400 bp. (**b**) DNA fragments less than 150 bp correspond to primer dimers or self-ligated product and must be removed before sequencing

4. Pool the three PCR reactions (75 µL) and purify DNA with SPRI magnetic beads.
5. Add 60 µL of beads (bead/DNA ratio of 0.8) and mix tenfold by pipetting (*see* **Note 15**).
6. Perform **steps 5–11** from Subheading 3.5.
7. Resuspend the beads with 20 µL of water.
8. Incubate at room temperature for at least 1 min.
9. Place the tube 2 min on the magnetic support and transfer supernatant (18 µL) to a new tube.
10. Check DNA purity, size, and concentration on Bioanalyzer (*see* **Note 16**, Fig. 3).

3.10 DNA Sequencing and Analysis

1. Mix barcoded samples ready for single-end or paired-end multiplexed sequencing with the Illumina HiSeq System.
2. No fewer than 30 million aligned reads are needed per sample to have sufficient depth and coverage of the human genome.
3. Check quality of raw sequence data obtained from high-throughput sequencing using a quality control tool such as

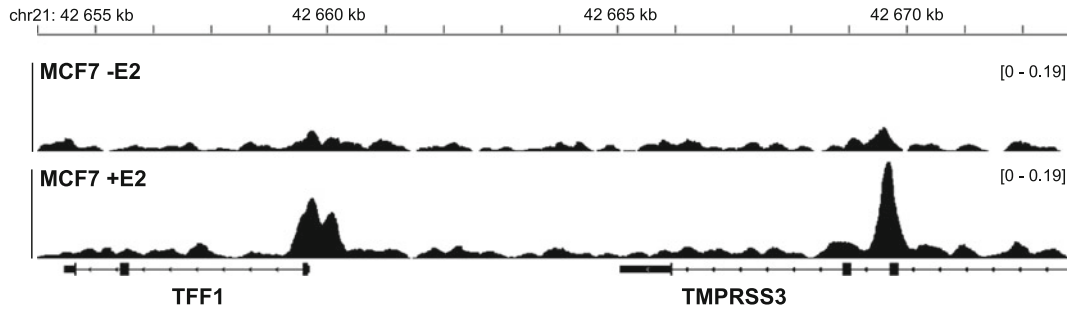


Fig. 4 Example of results obtained following sequencing. Aligned raw data were converted to Wig file and visualized with Integrative Genomics Viewer (IGV) [15]. Data were obtained from MCF7 cells treated with or without estradiol (E2) during 30 min

FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/) before further analysis.

4. Separate Illumina reads (demultiplexing) based on the sequence tag (barcode) using Novobarcode (Novocraft.com).
5. Align sequences on the genome reference using an algorithm such as Burrows-Wheeler Aligner (BWA) [11].
6. Verify mapping quality using SAMStat [12].
7. Aligned tags can be converted to WIG files using MACS [13] or F-Seq [14] and visualized with Integrative Genomics Viewer (IGV) [15] (*see* Fig. 4).
8. Perform peak calling using an algorithm such as MACS [13].

4 Notes

1. Sonication settings (i.e., intensity, time, and number of cycle) must be adjusted depending on the cell type and the apparatus. Bioruptor sonicator (Diagenode) was used in this protocol. For sonication with probe, increase the volume of cell lysis (500 µL) with dilution buffer (0.01 % SDS, 1.1 % Triton, 1.2 mM EDTA, 16.7 mM Tris pH 8.1, 167 mM NaCl, 1 mM PMSF, and 1× protease inhibitor cocktail before use). Prepare cells to test different sonication settings and check DNA fragmentation on agarose gel.
2. To verify sonication efficiently, take a small aliquot volume (5 %) of sonicated cell lysis and incubate samples with RNase A (20 µg) for 30 min at 37 °C, and then with 100 µg of proteinase K for 1 h at 65 ° C before load on a 1.2 % agarose gel. The ideal size of DNA fragments following sonication is between 200 and 500 bp (*see* Fig. 2).
3. Quick spin the phase-lock gels before adding the samples and phenol:chloroform.

4. Do not forget the input samples collected in **step 5** of Subheading 3.2.
5. Because the large amount of DNA obtained after FAIRE assay, elution can be done in 100–200 μL.
6. Before beginning library preparation, it is strongly advised to check FAIRE enrichment by qPCR analysis (FAIRE-qPCR) on specific regulatory elements such as transcriptional start site or enhancer and negative control region (*see* Fig. 1, right panel). The relative enrichment of specific open chromatin region can be calculated according to ΔCT method. The resulting threshold amplification levels of FAIRE samples (CT^{FAIRE}) are compared to the input DNA (CT^{input}) by subtraction ($CT^{FAIRE} - CT^{input} = \Delta CT$). For a relative DNA level, calculate $2^{-\Delta CT}$ for all samples. Input DNA can also be diluted to produce a standard curve that will give precisely the % input of the FAIRE enrichment.
7. SPRI beads are used to purify DNA from buffer reaction and to select fragments with specific size. Remove beads from 4 °C and warm to room temperature for at least 30 min before use. Because DNA recovery is strictly dependent on the volume ratio of SPRI bead suspension and DNA solution, it is very important to rigorously resuspend the beads by vortexing or by pipetting up and down. So precise pipetting performance is very critical for each step using SPRI beads.
8. Bead/DNA ratio of 1.0 allows fragment binding larger than 100 bp.
9. EtOH has hygroscopic property. It is very important to use freshly prepared 70 % EtOH for optimal results. Be careful not to disturb beads during wash.
10. Dry beads enough to remove all trace of EtOH before cracking appears on the beads to assure DNA elution efficiency.
11. This step is optimized for selection of 250–350 bp fragment size. Bead/DNA ratio of 0.7 allows fragment binding and discards fragments larger than 400 bp for the first step of selection.
12. Bead/DNA ratio of 0.15 allows fragment binding larger than 150 bp for the second step of selection.
13. Prepare annealed adapters. Resuspend lyophilized adapters to 100 μM in water. Mix the A adapters up and down together to a final concentration of 40 μM each in 10 mM Tris and 10 mM NaCl (do the same procedure with the B adapters). Anneal adapters using the following thermocycling program: decrease temperature by 1 °C/30 s from 98 to 4 °C. Then mix annealed adapters A and B together to a final concentration of 20 μM and store at −20 °C. For adapter ligation, use DNA/adapter ratio of 0.1 (1:10). Note the quantity and length of fragment for the calculation of adapter quantity to add:

$$x = \frac{a}{b \times 600\text{Da}} \times 10$$

x = adapter quantity need in nmol, a = DNA start quantity in ng (here 50 ng), b = average DNA fragment size selected (here 300 bp), and 600 Da = average molecular weight of a DNA base pair.

14. The number of cycles depends of the beginning quantity of DNA used and the success of the library preparation. Generally between 12 and 18 cycles are sufficient for DNA amplification.
15. Bead/DNA ratio of 0.8 allows removing more efficiently primer dimer following qPCR amplification.
16. To check DNA quality, use 1 μL diluted 1/5 or 1/10 and load on DNA chips for Bioanalyzer analysis (*see* Fig. 3).

Acknowledgement

We thank Alain Lavigueur and Maïka Jangal for critical comments on the manuscript.

References

1. Hogan C, Varga-Weisz P (2007) The regulation of ATP-dependent nucleosome remodelling factors. Mutat Res 618(1-2):41–51
2. Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. Cell Res 21(3):381–395
3. Jin J, Cai Y, Li B et al (2005) In and out: histone variant exchange in chromatin. Trends Biochem Sci 30(12):680–687
4. Cockerill PN (2011) Structure and function of active chromatin and DNase I hypersensitive sites. FEBS J 278(13):2182–2210
5. Rizzo JM, Sinha S (2014) Analyzing the global chromatin structure of keratinocytes by MNase-Seq. Methods Mol Biol 1195:49–59
6. Gaulton KJ, Nammo T, Pasquali L et al (2010) A map of open chromatin in human pancreatic islets. Nat Genet 42(3):255–259
7. Paul DS, Albers CA, Rendon A et al (2013) Maps of open chromatin highlight cell type-restricted patterns of regulatory sequence variation at hematological trait loci. Genome Res 23(7):1130–1141
8. Song L, Zhang Z, Grasfeder LL et al (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res 21(10):1757–1767
9. Waki H, Nakamura M, Yamauchi T et al (2011) Global mapping of cell type-specific open chromatin by FAIRE-seq reveals the regulatory role of the NFI family in adipocyte differentiation. PLoS Genet 7(10), e1002311
10. Rodrigue S, Materna AC, Timberlake SC et al (2010) Unlocking short read sequencing for metagenomics. PLoS One 5(7), e11840
11. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754–1760
12. Lassmann T, Hayashizaki Y, Daub CO (2011) SAMStat: monitoring biases in next generation sequencing data. Bioinformatics 27(1):130–131
13. Zhang Y, Liu T, Meyer CA et al (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9(9):R137
14. Boyle AP, Guinney J, Crawford GE et al (2008) F-Seq: a feature density estimator for high-throughput sequence tags. Bioinformatics 24(21):2537–2538
15. Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14(2): 178–192

Chapter 18

Aggregate and Heatmap Representations of Genome-Wide Localization Data Using VAP, a Versatile Aggregate Profiler

Mylène Brunelle, Charles Coulombe, Christian Poitras, Marc-Antoine Robert, Alexei Nordell Markovits, François Robert, and Pierre-Étienne Jacques

Abstract

In the analysis of experimental data corresponding to the signal enrichment of chromatin features such as histone modifications throughout the genome, it is often useful to represent the signal over known regions of interest, such as genes, using aggregate or individual profiles. In the present chapter, we describe and explain the best practices on how to generate such profiles as well as other usages of the versatile aggregate profiler (VAP) tool (Coulombe et al., Nucleic Acids Res 42:W485–W493, 2014), with a particular focus on the new functionalities introduced in version 1.1.0 of VAP.

Key words Data analysis, Graphical representation, Aggregate profiles, Average profiles, Heatmaps, ChIP-Seq

1 Introduction

Many whole-genome experiments are generating a quantitative measure of the localization of a chromatin feature of interest in cells through sequencing (e.g., ChIP-Seq, DNA-Seq, MNase-Seq; *see* [2] for review). These measures are usually enclosed in density files containing the signal throughout the genome at different resolutions and encoded in diverse file formats. The first obvious operation is to visualize the results in a genome browser (for instance, the UCSC Genome Browser [3] and the Integrative Genomics Viewer [4]). In cases where the chromatin feature of interest tends to have sharp enrichments such as transcription factors, peak-calling approaches are used to identify and characterize the localization of these enriched regions (*see* [5] for review). Other chromatin features such as certain histone modifications tend to produce more diffuse regions of enrichment that would often not fit the requirements of the peak-calling algorithms. In such cases, and even with

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_18,

features of sharp enrichment, it is useful to represent the signal intensity level over known biological annotations of the genome such as genes or other regions of interest. In some cases, the spatial distribution of the signal between different chromatin features is compared or the intensity level of the same feature in different conditions. The versatile aggregate profiler (VAP) running on both laptop and supercomputers with Linux, Mac OS X, or Windows was designed for that kind of analyses and is suitable for both experimental and computational biologists [1]. In VAP, the *Reference points* correspond to the anchor points on which the *Reference features* (e.g., genes) are aligned. For instance, one *Reference point* is used to generate a graph over the aligned transcription start sites (TSS), two *Reference points* are used to analyze the signal over genes (the two borders of the genes are aligned independently, but their orientation is considered), and up to six *Reference points* could be used as in the case of delimiting the signal over the first, middle, and last exons (*see* Fig. 2 of the original paper [1]). In the process, all the *Reference features* of a group are virtually aligned on the positive strand in order to represent all the 5′ boundaries to the left and all the 3′ to the right (*see* **Note 1**).

In this chapter, we describe the best practices to generate aggregate profiles as well as other functionalities using the new version 1.1.0 of VAP. Compared to the originally published version 1.0.0 [1], the version 1.1.0 described here includes (1) the support of the bigWig format for the *Datasets* through the use of code from the NGS++ library [6], (2) the support of the GTF format for the *Genome annotations* file, (3) the possibility to create a heatmap representation of the individual profiles, (4) the possibility to filter the *Reference groups* using genomic regions, (5) an improvement of the overall robustness, and importantly (6) the addition of a new parameter to process missing data in a *Dataset*. Indeed and as shown in Fig. 1, the inclusion of genomic regions without signal (called missing data) could have an important impact on the generation and interpretation of the aggregate graphs. The Subheading 2 of this chapter describes the software you have to install on your computer and the files that are required depending on the type of analyses you want to conduct. The Subheading 3 mostly describes and details typical analyses that are easily conducted using VAP.

2 Materials

2.1 Software

1. Download the complete VAP Java packaged file version 1.1.0 (http://lab-jacques.recherche.usherbrooke.ca/vap/downloads/). It contains all you need to run VAP on a laptop, desktop, or server computer with Unix/Linux, Mac, or Windows, including test files (*see* **Note 2**).

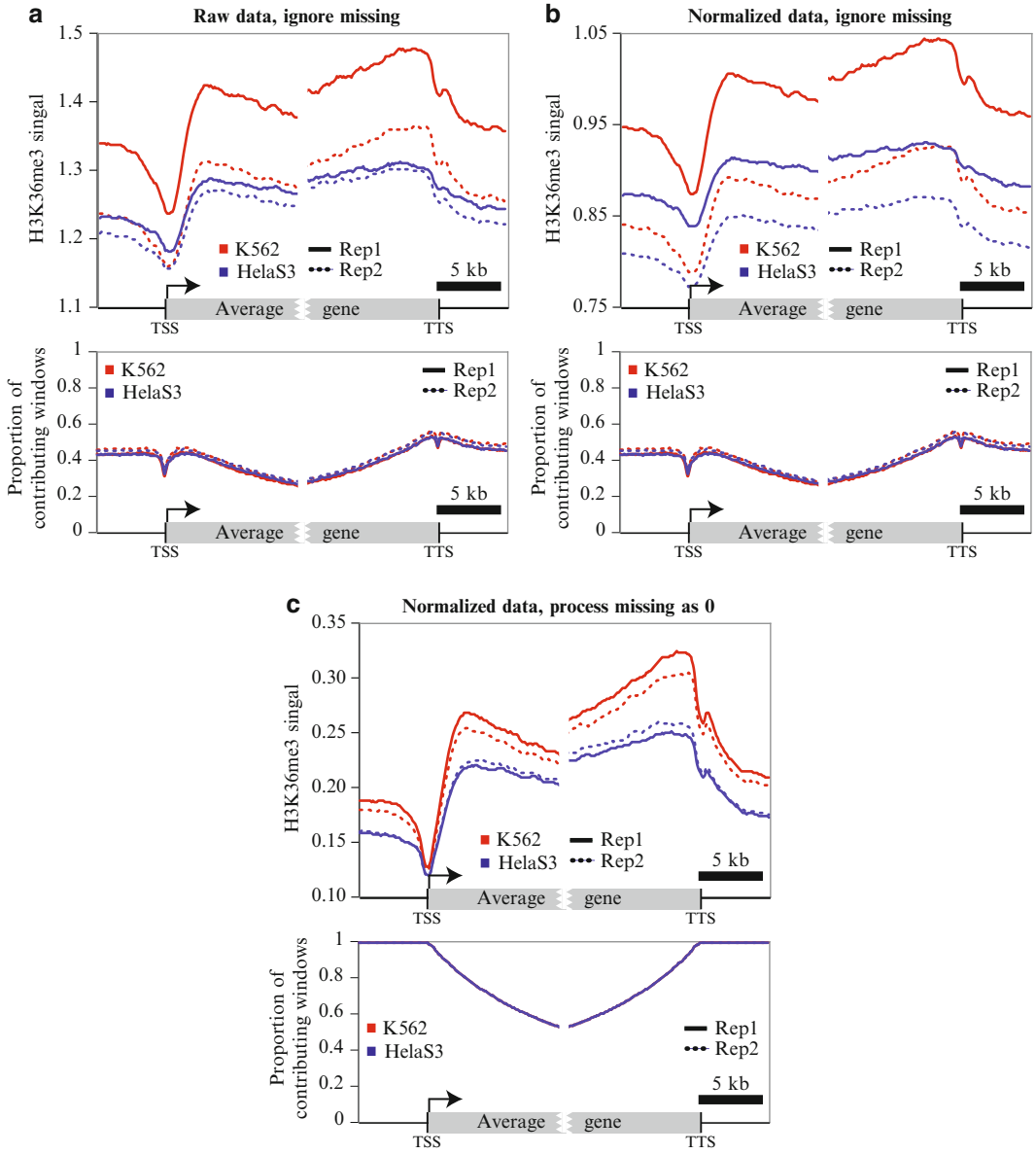


Fig. 1 Aggregate graphs generated with four normalized (or not) ChIP-Seq datasets of the histone modification H3K36me3 obtained from ENCODE (two biological replicates (*plain* and *dotted curves*) in two different cell lines (K562 (*red*) and HeLaS3 (*blue*))) and sequenced by UW [10], profiled on the 31,680 unique refSeq transcripts extracted from UCSC in September 2014. The BAM files were downloaded (from http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwHistone/) and used to generate the "raw" and "normalized" bigWig files (*see* **Notes 12** and **13**). The aggregate profiles were generated directly using the raw data and ignoring the missing data (**a**), or using the normalized data and ignoring (**b**) or processing (**c**) the missing data (*see* **Note 30** regarding the bottom "Proportion" graph of each panel). Based on the comparison of these three panels, it can be concluded that processing the missing data (**c**) allows the biological replicates to look much more similar than just normalizing for the number of mapped reads (**b**). All the files (including the parameters file) used to generate this Figure are available in lab's website (http://lab-jacques.recherche.usherbrooke.ca/vap/downloads/)

2. Download and install Java JRE 7.0 update 9 (or above) (https://www.java.com/download/) to open and use the VAP interface.
3. Make sure that everything is properly installed and configured by using the "Test VAP" menu. Briefly, this will select the *Annotations* analysis mode; load the genome-wide H3K36me3-vs-H3 [7] and H2A.Z-vs-H2B [8] yeast ChIP-chip *Datasets*; load three *Reference groups* corresponding to genes with high, mid, and low transcriptional frequency [9] as well as the corresponding *Genome annotations* file, and then generate one aggregate graph containing the two *Datasets* over the three *Reference groups* (for a total of six curves):
   (a) Open the interface by double-clicking on the Java archive (jar) file (*see* **Note 3**).
   (b) Click on the "Test VAP" menu of the interface, and then "Load test files."
   (c) Choose an output folder in the "Output selection" section.
   (d) Click on the "Run" button at the bottom left of the interface.
   (e) The results of the analysis will be generated in the selected output folder, including a file called "test_agg_graph_all.png" showing the aggregate profiles.
   (f) In case you got an error message, *see* **Note 4**.

2.2 Input Files

Depending on the type of reference the user wants to analyze (*Annotations*, *Exons*, or *Coordinates*), the required input files are not exactly the same (*see* Subheadings 3.2–3.4 for more details), but in all cases, the *Datasets* and *Reference groups* are mandatory. *See* **Note 5** about the coordinate intervals convention of all the input and output files (zero-based half-open vs. one-based closed).

1. *Datasets*: The *Datasets* are the density files summarizing the number of sequencing read aligned (or the hybridization intensity) at given position of the genome. The data formats currently supported in VAP for density files are bigWig, bedGraph, and WIG (https://genome.ucsc.edu/FAQ/FAQformat.html) (*see* **Note 6**). In a future version of VAP, we plan to also support aligned file formats such as BAM; meanwhile, if you are starting with such aligned files, please refer to Subheading 3.1 below.
2. *Reference groups*: The *Reference groups* are the groups of *Reference features* of interest (either genome annotations such as genes or user-provided regions such as peaks) on which you want to aggregate the *Dataset*'s signal. The content of a *Reference group* varies depending on the type of reference you want to analyze, but there is always one *Reference feature* per line.

In the *Annotations* and *Exons* analysis modes, a *Reference group* file contains only one column with the unique identifiers (e.g., gene names) written exactly (case sensitive) as in the corresponding *Genome annotations* file (*see* next step and **Note 7**). In the *Coordinates* analysis mode, the *Reference groups* directly contain the genomic coordinates (e.g., enriched regions for a particular transcription factor identified by peak-calling approaches) in a BED format (three to six columns) when there are exactly two *Reference points* or in a special format for analyses using one to six *Reference points* (*see* **Note 8**).

3. *Genome annotations*: This file is only required in the *Annotations* and *Exons* modes to extract the genomic coordinates of the *Reference features*. The formats supported are genePred and GTF (https://genome.ucsc.edu/FAQ/FAQformat.html) (*see* **Note 9**).
4. *Selection* or *Exclusion* filters: Optionally it is possible to select or exclude *Reference features* from the *Reference groups*, as well as genomic regions (*see* **Note 10**).

3 Methods

The Java interface proposes two main tabs that are self-explanatory: *Complete VAP process* and *Only create graphs*. Few sections of the interface are considered "advanced" and can be accessed by clicking on the "+" sign beside the information button. When the user clicks on the "Run" button, a "VAP_parameter.txt" file is automatically created by the interface in the output folder, which can be later imported through "Open parameters file" from the "File" menu to generate similar analyses (e.g., same input files but using different numbers of windows per block or a different windows size). The "File" menu also proposes a selection of parameters often used to analyze mammalian or yeast *Datasets* (the latter being much smaller and compact).

3.1 Prepare Density Files from Aligned (BAM) Files

This preliminary file-processing step is optional; in cases you already have density files, you can directly go to Subheading 3.2. As mentioned above, the BAM/SAM file format, which contains the information of each read aligned to the reference genome, is not yet supported by VAP and should thus be processed. If you are starting with such files, you should already have checked the quality of the raw sequences, applied a trimming step if relevant, mapped the reads, checked the quality of the mapping, and filtered the reads if required (*see* **Note 11**). Ideally you would generate normalized density files from aligned (BAM) files using a sophisticated method such as the one implemented in Wiggler [10, 11] that shifts the reads, combine biological replicates, normalize for

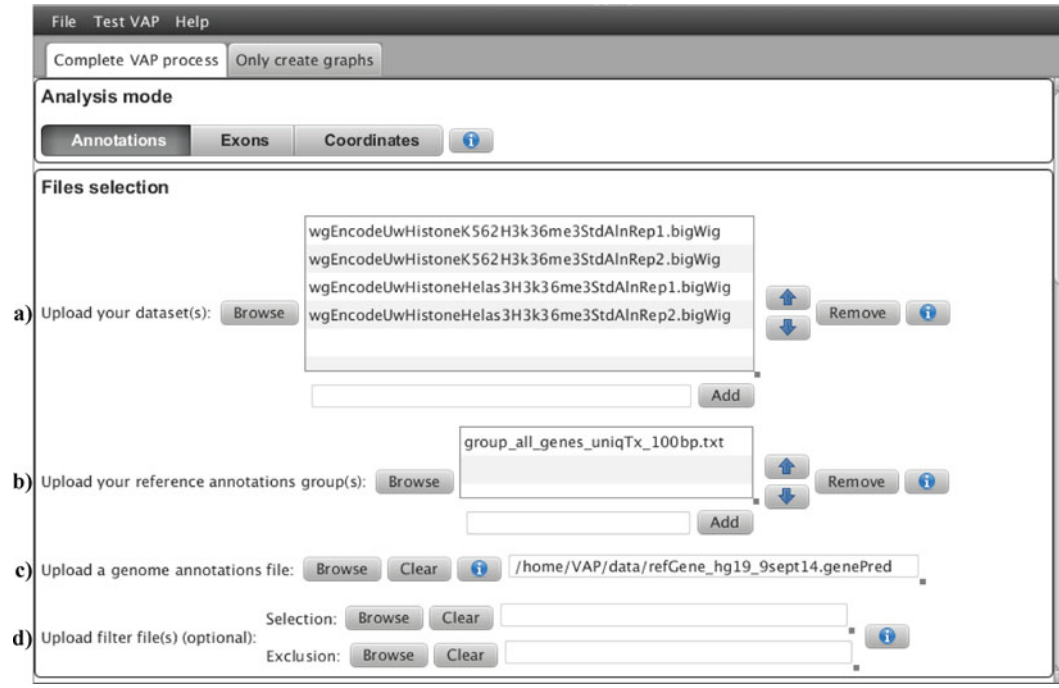


Fig. 2 Screenshot of the *Files selection* section showing the parameters described in Subheading 3.2 **step 1**

the number of mapped reads, and take into account the local mappability to distinguish missing values from unreliable genomic positions, in order to generate a *Dataset* containing the signal expressed in terms of a fold change over the expected signal from an equivalent uniform distribution of reads over all mappable locations in the genome (https://sites.google.com/site/anshulkundaje/projects/wiggler). Otherwise, minimally you want to:

1. Create a density file (bedGraph format) where the signal is normalized for the number of mapped reads (*see* **Note 12**).
2. Convert this bedGraph file into a much more compressed bigWig file (*see* **Note 13**).

3.2 Generate Aggregate Profiles over Genes (or Other Annotations from a Genome Annotations File)

This is the most popular analysis mode of VAP, accessible from the main *Complete VAP process* tab of the interface.

1. In the *Files selection* section (*see* Fig. 2), drag, browse, or write the name of:
   (a) The normalized *Dataset*(s) (generated at Subheading 3.1 or obtained externally) in one of the currently supported density format (*see* Subheading 2.2 **item 1**). The bigWig format is highly recommended because it is much more efficient to process (*see* **Note 14**).
   (b) The *Reference group*(s) (*see* Subheading 2.2 **item 2** and **Note 15**).

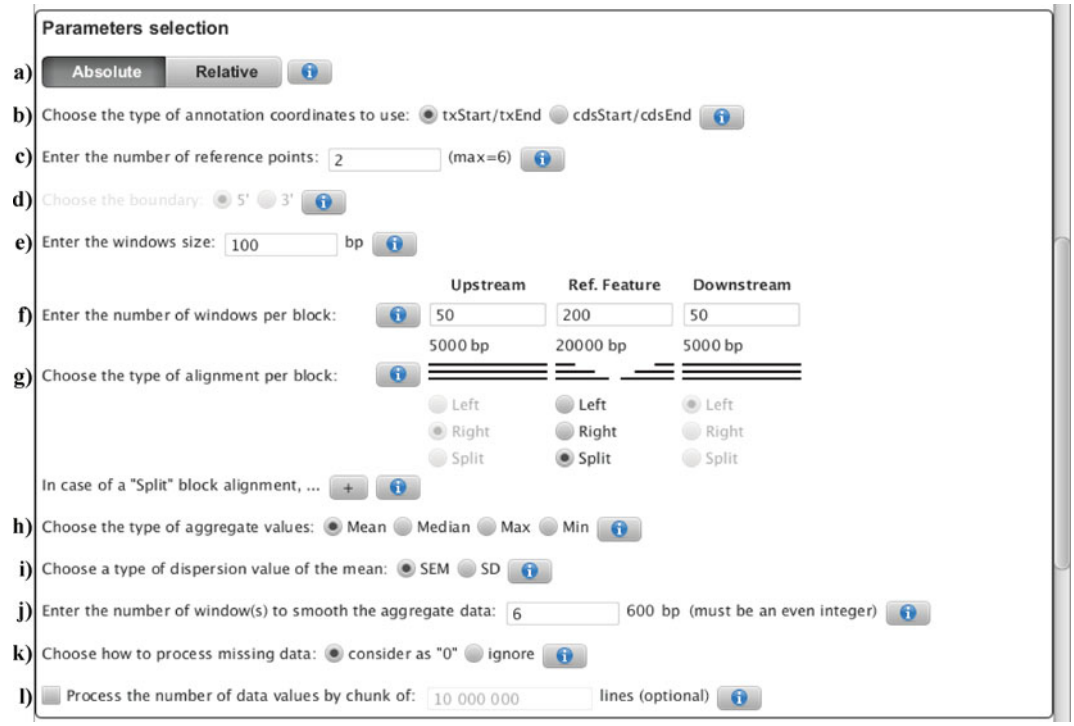


Fig. 3 Screenshot of the *Parameters selection* section showing the parameters described in Subheading 3.2 **step 2**

(c) The *Genome annotations* file (*see* Subheading 2.2 **item 3**).

(d) Optionally you can use *Selection* and *Exclusion* filter file(s) (*see* Subheading 2.2 **item 4**).

2. In the *Parameters selection* section (*see* Fig. 3), you have to select or enter:

(a) The analysis method using the toggle buttons. The *Absolute* method uses windows of constant size (meaning that longer feature contains more windows) per block, while the *Relative* method uses a constant number of windows (meaning that longer feature contains longer windows). Considering that the *Absolute* method is almost always recommended (*see* **Note 16**), only the parameters related to this method are described here.

(b) The type of annotation coordinates (either transcriptional or coding) to extract from the *Genome annotations* file.

(c) The number of *Reference points* (*see* **Note 17**). By definition, the number of blocks is the number of *Reference points*+1 (one block each side of each *Reference point*).

(d) The 5′ or 3′ boundary in the case where only one *Reference point* is selected.

(e) The size of the windows used to segment each block (and by extension the *Reference features*) (*see* **Note 18**).

(f) The number of windows used in the representation of each block individually (*see* **Note 19**).

(g) The type of alignment per block (*see* **Note 20**).

(h) The type of aggregate values to report for each window of the *Reference groups* (*see* **Note 21**).

(i) The type of dispersion value in the case where you selected to report the mean as the aggregate values (*see* **Note 22**).

(j) The number of consecutive windows to average in order to smooth the aggregate data (*see* **Note 23**).

(k) How to process missing data. This new option of VAP is particularly important to get more accurate aggregate profiles in cases where the reliable genomic regions without signal are not included with a value of zero in your *Datasets* (*see* Subheading 3.1). As mentioned by others (https://sites.google.com/site/anshulkundaje/projects/wiggler), including these regions with zeros in the analysis is critical (*see* Fig. 1).

(l) Optionally you can control the memory footprint of VAP by deciding to process a limited number of datavalues at a time.

3. In the *Output selection* section (*see* Fig. 4), you have to select or enter:

   (a) The output folder where all the output and related files will be generated.

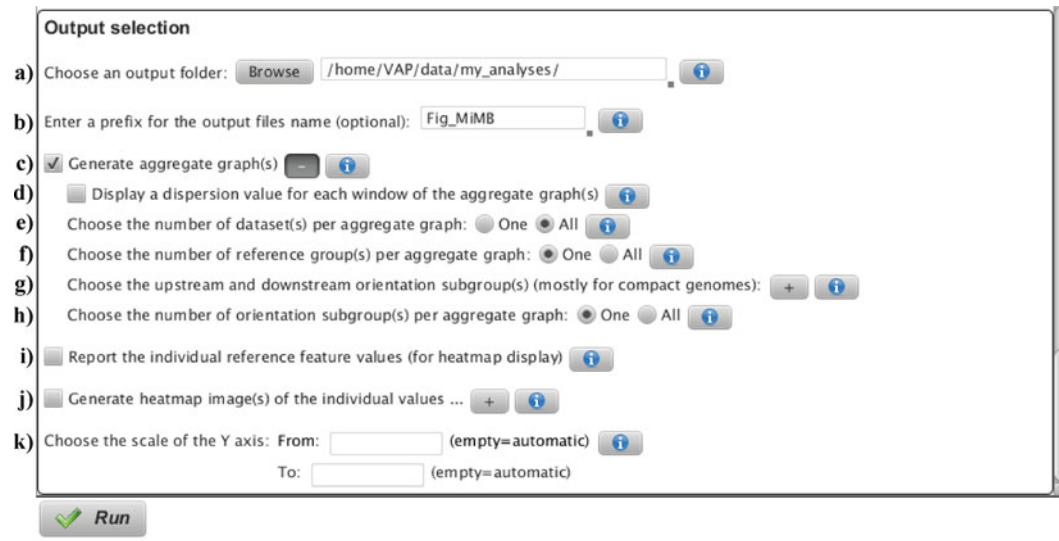


Fig. 4 Screenshot of the *Output selection* section showing the parameters described in Subheading 3.2 **step 3**

(b) Optionally you can add a prefix that will be added to all the output file names (*see* **Note 24**).

(c) Optionally you can choose to generate the aggregate graph(s).

(d) Optionally you can choose to display a dispersion value at each window of the aggregate graph (selected at **step 2i**).

(e) The number of *Dataset*(s) per aggregate graph (one or all).

(f) The number of *Reference group*(s) per aggregate graph (one or all).

(g) Optionally you can choose to automatically subgroup the *Reference features* based on the orientation of the upstream and downstream *Genome annotations* (*see* **Note 25**).

(h) The number of orientation group(s) per aggregate graph (one or all).

(i) Optionally you can choose to output the individual *Reference features* values. This will generate one file per *Reference group* and per *Dataset*, mainly containing one individual *Reference feature* per line and the signal of individual windows in the columns. Considering this file is essential for heatmap representation, this box is automatically checked if the next option is also checked.

(j) Optionally, you can choose to create heatmap image(s) of the individual *Reference features* values with predefined or customized settings (*see* **Note 26**).

(k) Optionally you can define the scale of the Y-axis; otherwise, it will be populated automatically (*see* **Note 27**).

Note that at least one of the aggregate, individual, or heatmap output has to be selected.

4. Click on the green "Run" button (*see* **Note 28**).

(a) The interface validates the parameters you entered; in case of an error, a pop-up window appears.

(b) When everything is fine, a progression bar appears, indicating the name and the proportion of the total number of steps to accomplish.

(c) When the execution is completed, a new window appears containing either the "Analysis completed" message, or the error message(s) in case of a problem.

5. Interpretation of the output files.

There are four groups of output files (*see* **Note 29** for the naming scheme):

(a) Aggregate graphs: When this option is selected, VAP generates png images (using the sub-prefix "agg_graph") containing two plots; the first contains the aggregate values in

each window of each block (e.g., upstream region, *Reference feature*, and downstream region when using two *Reference points*), while the second contains the corresponding information on the proportion of the *Reference features* contributing to the aggregate value (*see* **Note 30**). To generate these images, other text files are produced containing the actual values (using the sub-prefix "agg_data") as well as the list of aggregate graph to generate (using the sub-prefix "list_agg_graph") (*see* **Note 31**). These text files can also be used by macro included into Libre Office and Microsoft Office spreadsheets available in the download section of VAP to generate vectorial graphs that can be imported to image editing software such as Illustrator or Inkscape to generate publication-ready figures. Importantly, you have to be careful not to over-interpret the information in these graphs (*see* **Note 32**).

(b) Heatmap representation: When this option is selected, VAP generates png images (using the sub-prefix "heatmap") containing the individual values of each window of each block of each *Reference feature*. To generate these images, other text files are produced containing the actual values (using the sub-prefix "ind_data") as well as the list of heatmap to generate (using the sub-prefix "list_heatmap") (*see* **Note 33**).

(c) Parameters file: This file contains one parameter per line in a special format (*see* **Note 34**). It is using the sub-prefix "VAP_parameters.txt" when created by the interface, or any other name when created or modified by the user.

(d) Log files: Many execution details (including warning and error messages) are written in both the vap_core and vap_interface log files.

3.3 Generate Aggregate Profiles over Exon Boundaries (Exons Mode)

This analysis mode is probably the less popular but nonetheless very useful in certain cases. Since this mode is almost identical to the *Annotations* mode (*see* Subheading 3.2), only the different parameters are enumerated:

1. In the "Parameters selection" section:

   (a) The coordinates used are necessarily those from transcript boundaries; therefore, the parameter from Subheading 3.2 **step 2b** is not displayed.

   (b) The number of *Reference points* at Subheading 3.2 **step 2c** is automatically six, corresponding to the beginning and end of all the first, middle, and last exons, and consequently introns (*see* **Note 35**).

(c) An additional parameter allows to merge (or not) the signal of the middle intron(s) (for genes having more than three exons) with the first or last introns.

2. In the "Output selection" section, the two parameters of the aggregate graphs regarding the orientation subgroup(s) (Subheading 3.2 **steps 1g–h**) are not relevant in this analysis mode and are therefore not displayed.

3.4 Generate Aggregate Profiles over User-Defined Regions (Coordinates Mode)

Since this analysis mode is almost identical to the *Annotations* mode (*see* Subheading 3.2) (the biggest difference is the content of some files (*see* Subheading 2.2)), only the different parameters are enumerated:

1. In the "Files selection" section, since the coordinates of the *Reference points* are included in the *Reference group*, no *Genome annotations* file (Subheading 3.2 **step 1c**) is therefore required.
2. In the "Parameters selection" section, the type of annotation parameter (Subheading 3.2 **step 2b**) is not relevant since only one type is provided in the *Reference group*.
3. In the "Output selection" section, the two parameters of the aggregate graphs regarding the orientation subgroup(s) (Subheading 3.2 **steps 1g–h**) are not relevant in this analysis mode and are therefore not displayed.

3.5 Generate Individual Average Value per Gene or Regions

Even if there is no specific analysis mode for this application in the interface, it is worth to note that it is possible to choose parameters allowing calculating the average signal value of each gene or region of interest. Follow the method of Subheading 3.2, except:

1. In the "Parameters selection" section:
   (a) Choose a very large windows size (e.g., bigger than the longest *Reference feature*).
   (b) Choose two *Reference points*.
   (c) Choose only one window per block.
2. In the "Output selection" section, only check the option to report the individual *Reference feature* values (the aggregate and heatmap graphs are not relevant in such a case).
3. VAP will generate text file(s) named "ind_data" containing three columns of data (*see* **Note 33**) where the middle one corresponds to the average signal over each *Reference feature* (the first and third columns of the data correspond to the upstream and downstream regions, respectively). It is then possible to sort the *Reference features* based on this value to create new *Reference groups* of similar level of enrichment and use these new groups to generate aggregate profiles. *See* **Note 36** for alternative usages.

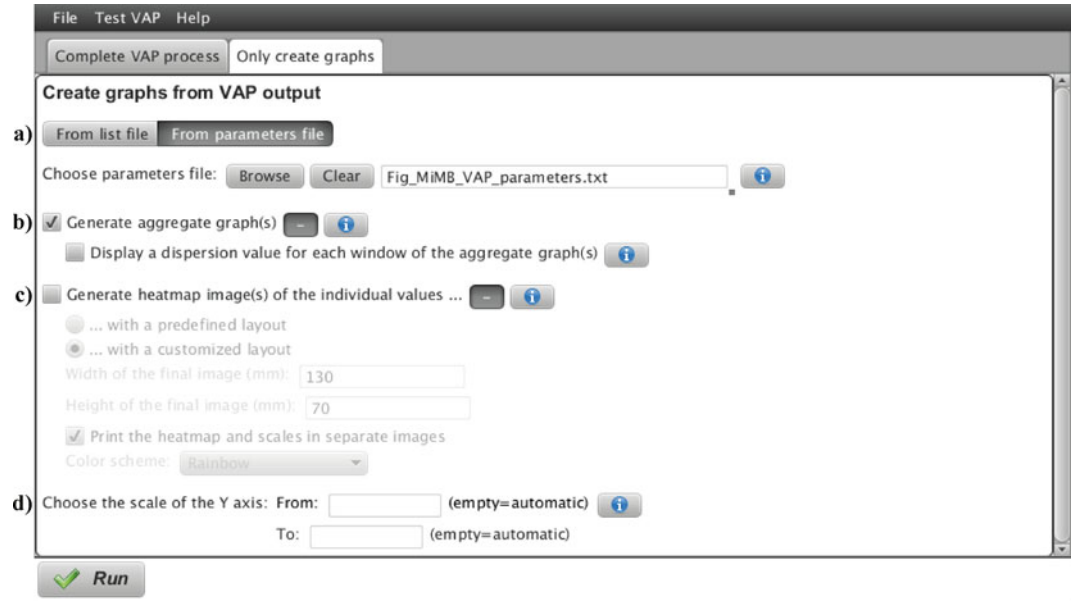


Fig. 5 Screenshot of the *Only create graphs* tab of the interface showing the parameters described in Subheading 3.6 **step 1**

3.6 Only Create Graphs

In some cases, it is desirable to generate the graphical representations of an analysis without having to rerun it. Two such frequent scenarios are (a) you want to change the content (*Datasets*, groups, order) of an aggregate or heatmap graph by simply manually modifying (remove, replace) the content of a list file (*see* Subheading 3.2 **steps 5a** and **b**); (b) you directly ran the vap_core module from a terminal (*see* **Note 37**), therefore without using the vap_interface module that is the one generating the graphs. There are two different ways to only generate the graphs:

1. Use the "Only create graphs" tab of the interface (*see* Fig. 5):
   (a) You can either choose to use the "From list file" or "From parameters file" toggle button and then select the desired aggregate or heatmap graph(s) and their appropriate available parameters. Using the list file(s) does not allow drawing all the elements of the *X*-axis (*see* **Note 38**).
   (b) You can choose to display (or not) the dispersion values on the aggregate graph(s).
   (c) You can choose to create (or not) heatmap image(s) with predefined or customized settings (*see* **Note 26**).
   (d) You can define the scale of the Y-axis; otherwise, it will be populated automatically.
2. Run the vap_interface module from the terminal where you can define the same parameters than from the interface through the use of different options (*see* **Note 39**).

4 Notes

1. Consequently, a genomic region containing a *Reference feature* on the positive strand (P) is used as is, while the region containing a *Reference feature* on the negative strand (N) and its upstream and downstream annotations is inverted. For instance, a region containing the genes A-B-C (ordered from their increasing genomic coordinates) where genes A and B are on the negative strand and gene C is on the positive strand is therefore on native orientations NNP; in a case where gene B is included in a group of *Reference features*, the whole region is virtually inverted such that this region is used as if the genes were ordered C-B-A and their orientations were NPP.
2. More precisely, the complete VAP tool is a Java archive (jar) file corresponding to the `vap_interface` module packaged with the C++ `vap_core` binaries compiled statically (therefore no dependency) for the supported OS (Unix (most distributions), Mac OS X, Windows XP, and above; 32 and 64 bits architectures). The individual modules can be optionally downloaded separately.
3. Alternatively, type in a terminal `java -jar vap-1.1.0-all.jar` from the folder containing the jar file.
4. It is preferable to avoid space and particularly accent in all the paths and file names (program, files, output directory). Since the two log files generated by VAP are simple text files, do not hesitate to take a look inside, and if you still do not understand the error, please contact vap_support at usherbrooke.ca by providing the two log files with an explanation of what you are trying to do. Eventually we might ask for your files to reproduce the error (*Reference groups*, *Genome annotations*, *Datasets*).
5. In version 1.1.0, a small bug was corrected to make sure to correctly interpret the coordinates of all the input file formats as defined by the UCSC Genome Bioinformatics team. This means that for genePred, BED, bedGraph, and bigWig formats, the first position of the chromosome is "0" (zero-based) and the start of an interval is included, but the end is not (half-open). For example, in this convention, the interval `chr1 1 10` contains nine positions (1–9) starting at the second position of chromosome 1. Since the format definition of GTF is one-based and uses closed intervals (the end position is included in the interval), it is therefore normal that the same gene represented in genePred and GTF format will apparently not have the same coordinates, but the results of VAP will be exactly the same. The WIG data format is also one-based closed. *See* **Note 8** about the special format required in the *Coordinates* analysis mode. Internally, all the coordinates are represented as zero-based closed (therefore inclusive) intervals.

The coordinates in the "agg_data" output files are zero-based (in the way that 0 represents the first position of the *Reference features*), while the coordinates in the "ind_data" files follow the same convention than the corresponding *Reference group* file.

6. The file extension is first used to determine the file format, and then the attribute "type" (which is mandatory for *Datasets* and *Coordinates* files) from the definition line (first line of the file) is used to cross-validate the format (https://genome.ucsc.edu/FAQ/FAQformat.html). The supported extensions (case insensitive) are bigWig, bw, bedGraph, bg, wiggle, and WIG for *Datasets*; genePred, gp, and gtf for *Genome annotations*; and bed and coordX (where *X* is a value between 1 and 6; *see* **Note 8**) for *Coordinates* files. The extension is not used for the other *Reference group* files (in *Annotations* and *Exons* mode) and *Filter* files. If the extension is unknown or the attribute "type" does not correspond to the extension (not cross-validated), the file is skipped, and a warning is written in the vap_core log file. Note that the *Datasets* are limited to one track of data per file (otherwise the datavalues are all appended to the first track). Also note that in all *Dataset* and *Reference group* files where the definition line contains the attribute "name," this information is used in the name of the output files; otherwise, the file name is used.
7. If a *Reference group* contains the same *Reference feature* more than once, all instances are kept (and linked to the same genome annotation in *Annotations* or *Exons* mode). Optionally, the definition line can contain the attribute `#name=<file_alias>` where the text must be enclosed in quotes if it contains spaces. Any other line starting by "#" is considered as a comment. Note that only the first column is read (others ignored).
8. The first line of a *Reference group* file in the *Coordinates* mode (also directly called coordinates file) should start by a "#" and contains at least the attribute "type=" followed by one of the six "coordX" where *X* is the number of *Reference points* contained in this file. The full description of this definition line is `#type=coord<X> [name=<file_alias> desc=<file_description>]` where the three attributes can appear in any order and where the text of the name and description attributes must be enclosed in quotes if it contains spaces. Any other line starting by "#" is considered as a comment. Because this type of file can contain up to six *Reference points*, the order of the columns is slightly different than in the traditional BED format: the first column of the other lines must contain the chromosome, the second column, the strand ("+" or "−" (any other character interpreted as "+")), and the *X* following columns must contain the coordinates. The columns must be delimited by a tab or space characters (multiple instances are

processed as one, such that empty columns are not allowed). Note that in cases where a line contains less than *X* columns, the default values (which are an empty string for the chromosome, a dot for the strand, and −1 for the coordinates) are used for the missing columns. An additional column (*X*+1) can optionally contain an *alias* (e.g., unique identifier) that is then reported in the individual output file (*see* Subheading 3.2 **step 5c**). The *name* of such *Coordinates Reference feature* is always defined as the concatenation of all the required columns, delimited by a "_" (e.g., chr1_+_181135) that can be used as a *name* in the *Filter* files (*see* **Note 10**). Because most of the input files are zero-based (*see* **Note 5**), we decided to also use the zero-based convention for the coordinates contained in this file, and they are used as is (therefore the equivalent of closed intervals). To illustrate this special format, these are the first three lines of a coordinates file containing one *Reference point*:

```
#type=coord1 name=tRNA desc="tRNA trx start,
id=original tRNA coordinates"

chr1 + 181135 tL(CAA)A_181135_181248

chr1 - 139256 tP(UGG)A_139154_139256
```

And the first lines of a coordinates file containing four *Reference points* are the following:

```
#type=coord4 name="unannot transcripts"
desc="w surrounding genes"

chrI + 401 500 1501 1600 end-geneA_unannot1_
start-geneB

chrII - 1551 1650 2651 2750 end-geneG_
unannot2_start-geneH
```

Because we want the *Coordinates* analysis mode to be able to replicate the results obtained in the *Annotations* mode, the coordinates are associated to the blocks the same way that the annotation coordinates (start and end) are used in the *Annotations* mode (*see* **Note 17**), except that by using one *Reference point*, we assume it represents the equivalent of the 5′ of the annotation. If you are not sure, to understand correctly, look at the interface, and every time that a block is named "Inter," the coordinates of this block are derived from the coordinates of the neighbor blocks (using the +1 and −1 operations described in **Note 17**). Note that the end coordinate of the block named "Upstream" as well as the start of the block "Downstream" is derived if they are a neighbor of the block named "Reference feature"; otherwise, they use the coordinates provided in the coordinates file (as the flanking genes in *Annotations* mode). Considering that *Genome*

annotations files are either zero-based half-open (genePred) or one-based closed (GTF), we now provide a tool (included in a suite of utilities called vap_utils, http://lab-jacques.recherche.usherbrooke.ca/vap/downloads/) facilitating the conversion of a *Genome annotations* file to a coordinates file.

9. If you are working with an organism supported by the UCSC Genome Browser, it is possible to extract the *Genome annotations* file in genePred format from the "Table Browser" (http://genome.ucsc.edu/cgi-bin/hgTables) by selecting the option "selected fields from primary and related table" from the "output format," and then "check all" from the primary table except the first "bin" line. You can also download a *Genome annotations* file in GTF format from there. In genePred format, exon coordinates must be right flanked by a comma (e.g., "200,500,") and contain at least ten columns. When an 11th column is present, it is considered to be the *alias*, and this information is used as the second column in the "ind_data" output file, while other columns (12th and +) are copied at the end of the lines. In GTF format, only the "exon," "CDS," and "stop_codon" features (third column) are considered to reconstruct the transcripts, and the "transcript_id" attribute is used as annotation *name* while the "gene_id" attribute is used as *alias*. All *Genome annotations* files are represented internally as genePred format; therefore, non-coding annotations (lacking "CDS" feature) are using the end of the last exon coordinate as CDS coordinates (both for the start and end). If the *Genome annotations* file contains the same annotation *name* (supposedly unique identifier) more than once, the corresponding *Reference feature* of the *Reference group* is linked with the first instance of the *Genome annotations* file. Note that a header line(s) can be present but should start by "#."
10. The *Selection* filter is applied before the *Exclusion* filter. The *Filter* files can contain a mix of annotation names (unique identifiers in one column) and genomic coordinates (expressed in the three column tab-delimited BED format (`chr start end`) therefore using the zero-based half-open convention (*see* **Note 5**)). Note that for genomic coordinates, an overlap of 1 base pair with the *Reference feature* is sufficient to apply the filter, and exceeding the size of the chromosome is tolerated (e.g., if you don't know the exact size of a chromosome to exclude, use a very big number as end coordinate). For example, one of your *Reference group* contains all the transcribed genes in a given condition, but you want to restrict your analysis to the genes longer than 1 kb (the *Selection* filter therefore contains the list of long gene names) not in mitochondrial chromosome as well as not overlapping another annotation (the *Exclusion* filter therefore contains in the same file a line with something like `chrM 0 50000000` and the list of overlapping gene names). In the *Coordinates* mode, the filters represented as genomic

coordinates are applied on the columns corresponding to the *Reference feature* coordinates (*see* **Notes 8** and **17**).

11. Without going in details because it is beyond the topic of this chapter, some of the tools that can be used to conduct these quality control and preprocessing steps include FastQc (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/), Trimmomatic [12], BWA [13] or Bowtie [14], SAMtools (flagstat and view) [15], SAMStat [16], and Picard (https://github.com/broadinstitute/picard).

12. If you are comfortable to work from the terminal, here is an example bash script using SAMtools [15] and BEDtools [17] to generate normalized bedGraph files from already filtered BAM files (using the number of mapped reads for the normalization):

```
Files=*.bam
for i in $Files
do
    nbReadsMap=$(samtools view -F 4 -c $i)
    sclFctr=$(echo "scale=5; 10000000/$nbReads-
    Map" | bc)
    bedtools genomecov -bg -scale $sclFctr -ibam
    $i -g chromsize > $i-scl.bedgraph
    done
```

    where the file `chromsize` is tab-delimited and contains the name and the size of each chromosome on two columns and can be obtained from the UCSC Genome Browser. Note that using the `-bga` option of `genomecov` rather than `-bg` generates bedGraph files that include the regions with zero signal (giving much bigger files). However, given the new parameter of VAP to process missing data, you do not need to use this `-bga` option anymore to obtain unbiased profiles (*see* Fig. 1).

    If you are not comfortable with the terminal, it is possible to use SAMtools and BEDtools from the public web-based Galaxy platform (https://usegalaxy.org) [18] to generate normalized bedGraph files.

13. From the terminal, the easiest way is to use the `bedGraphToBigWig` utility from Kent's tool (http://hgdownload.cse.ucsc.edu/admin/exe/) by adding the following line to the script of **Note 12**:

```
./bedGraphToBigWig $i-scl.bedgraph chromsize $i-scl.bigwig
```

    This tool is also available from the Galaxy platform.

14. The reason why the bigWig format is much more efficient to process is that we do not have to read the whole file but only the relevant sections required to accomplish the requested analysis.
15. VAP can process more than one round of *Datasets* and *Reference groups* using the same parameters (e.g., generate the profile of *Datasets* A and B on the 500 most expressed genes in condition 1 and *Datasets* C and D on a different group of the 500 most expressed genes in condition 2). This can be done by manually modifying the parameters file to use different "alias" for the "dataset_path" and "refgroup_path" parameters. Unfortunately this feature is not available for now through the interface; you should rather use the vap_core module directly.
16. We designed VAP to use the *Absolute* analysis method by default because it is consistent with most of the molecular biological processes, but we nonetheless incorporated the *Relative* method since it is still used in the literature. For instance, the transcriptional processes can sense the absolute distance from a certain point (such as the TSS) based on the suite of posttranslational modifications involved at the different steps of transcription and always following the same order; the combination of marks at a certain point can therefore represent a "ruler" of the absolute distance [19]. In *Absolute* mode, each feature is divided in windows of constant size, while in *Relative* mode each feature is divided in a constant number of windows (therefore having varying length for two features of different length). The *Relative* mode implies that a signal appearing at a certain absolute distance from a point of reference (e.g., the H3K36me3 histone modification appearing after few hundred base pairs from the TSS of each gene) is not represented in the same window for short vs. long genes (e.g., a signal 600 bp downstream of the TSS is contained in the sixth window of a 1 kb gene divided in ten windows, but in the second window of a 3 kb gene, it's also divided in ten windows).
17. To help to understand the following explanations, change the number of *Reference points* on the interface, and you will notice an update of the layout. By using two *Reference points*, three blocks are generated to isolate in a block the *Reference feature* (often a gene) in the middle, while the upstream and downstream regions contain signal from flanking intergenic regions (potentially contaminated by signal from flanking genes depending of the length of the regions and the blocks). To completely delimit the flanking intergenic regions and avoid contamination from flanking genes, four *Reference points* are required (five blocks); in such a case, the upstream and downstream regions contain (but are not delimited by) signal from

the flanking genes. To completely delimit these flanking genes in separate blocks, six *Reference points* must be used (for a total of seven blocks). Using three and five *Reference points* allow to delimit the 5′ regions.

In more details, here is the description on how the coordinates of the *Reference features* (and its flanking annotations) are associated to the blocks: with one *Reference point*, if the 5′ boundary is selected (*see* Subheading 3.2 **step 2d**), the coordinate corresponds to the start of the second block (named "Downstream"), while if the 3′ boundary is selected, the coordinate corresponds to the end of the first block (named "upstream") in order to avoid a shift of one base pair compared with the results of two *Reference points* where the two coordinates of the *Reference feature* are used inclusively as the start and end of the middle block (named *Reference feature*); with three *Reference points*, the end coordinate of the upstream annotation corresponds to the end of the first block (named "Upstream"), while the two coordinates of the *Reference feature* correspond to the start and end of the third block (named "Reference feature") in order to isolate in the second block (named "Inter") the signal of the upstream intergenic region (the coordinates of this second block are derived by applying +1 and −1, respectively, on the end and start of the neighbor blocks); four *Reference points* are a derivative of the three *Reference points* where the start of the last block (named "Downstream") corresponds to the start of the downstream annotation to isolate the signal of the downstream intergenic region in the forth block; five *Reference points* are a derivative of the four *Reference points* where the signal of the upstream annotation is completely isolated in the second block (named "Annotation"); lastly, six *Reference points* are a derivative of the five *Reference points* where the sixth block (also named "Annotation") uses the coordinates of the downstream annotation.

18. This number represents the resolution of the output. In each window, the calculated value is always the weighted average, the weight being the coverage (footprint) of the datavalues (e.g., in a 100 bp window where there are two datavalues, respectively, with a value of 1.0 and a coverage of 75 bp, and with a value of 3.0 and a coverage of 25 bp, the weighted average is 1.5). In VAP version 1.0.0, the calculated value per window was the non-weighted average (e.g., given a result of 2.0 in the previous example), explaining that the results between the original and new VAP versions could be slightly different.

19. In the *Absolute* mode, you have to choose the length of each regions you want to represent in the aggregate or individual profiles and therefore use for each *Reference feature* and its neighbor regions; this length is simply calculated as the product of the windows size times the number of windows in a block. A gene shorter than the block capacity is completely included in the aggregate profile, meaning that some windows of the block are empty for this *Reference feature* (e.g., to represent a 2500 bp gene in a block having a capacity of 3000 bp (60 windows of 50 bp), there are ten windows around the split point (*see* **Note 20**) that stay empty). Conversely, the portion around the split point of a gene longer than the block capacity is not represented in the aggregate profile (e.g., the middle 1 kb of a 4 kb gene is not represented in the same block of a 3 kb capacity). It is recommended in the exploratory phase of your data analysis to select a number of windows per block that corresponds to approximately twice the average length of the regions contained in this block and eventually reduce this number to cover the average length. For instance, in a compact genome such as yeast, considering that the average gene length is ~1500 bp and the average intergenic region is ~500 bp, it is recommended to first use four *Reference points*, a windows size of 50 bp, and a number of windows per block of 10, 20, 60, 20, and 10 (completely covering regions of 500, 1,000, 3,000, 1,000, 500 bp, respectively).
20. In *Absolute* mode, the content of each block can be totally aligned to the left or to the right or split to align separately both the starts and the ends of the *Reference features*. Because the genomic coordinates of each side of a split point are not necessarily contiguous (e.g., gene longer than the block capacity; *see* **Note 19**), a gap is always introduced at the split point inside a block. Note that the split point is the end (100 %) or the beginning (0 %) when the block is aligned to the left or the right, respectively. Using the advanced parameters, the split point can be expressed in absolute or proportional distance of a boundary. Note that the alignment of the first (upstream) and last (downstream) block cannot be split because all the genomic regions inside these blocks have the same length, corresponding to the capacity of the block.
21. This parameter indicates to report one of the supported aggregate values (mean, median, max, or min), calculated across corresponding nonempty windows of all the *Reference features* of a given *Reference group*. This value must not be confused with the weighted average calculated from the datavalues for each window of each *Reference feature* mentioned in **Note 18**.
22. When the type of aggregate value is not the mean, no dispersion value is reported because it would be incorrect to show in a graph one of the other possible aggregate value (median,

max, or min) with the dispersion of the mean (standard deviation or standard error of the mean).

23. This value must be an even integer because the smoothed value reported at a given position is the result of the average calculated on the given window and half of the selected number of windows each side of it. This procedure is therefore creating a gap of the same size (half of the selected number of windows) at the split point of each block, represented as invalid windows. It is important to note that this smoothing operation is calculated only on the aggregate values, therefore not affecting the individual *Reference features* data reported in the "ind_data" files. Also note that for the windows positioned at less than half of the selected number of windows of the end of the block, the smoothed value is still calculated but consequently using less values (ignoring the neighbor invalid windows).
24. It is sometimes more useful to centralize all the output in the same folder with different prefixes than to create a different output folder each time, particularly in the exploratory analysis phase. Note that the only output files on which the prefix is not applied are the log files.
25. This option could be really important to correctly interpret signal in proximal regions that could be affected by the orientation of the flanking genes in compact genomes. It is only available in the *Annotations* analysis mode because we cannot assume that all the relevant regions of a genome are included in the *Coordinates* mode, and it is unlikely that the orientation of the neighbor genes can have an impact on the signal in the *Exons* mode. By default, the only subgroup selected is the one containing all the features, but it is possible to select subgroups based on the orientation of the upstream and downstream annotations of the *Reference feature*, named using the following nomenclature: A stands for *Any*, C stands for *Convergent* (tail-tail), D stands for *Divergent* (head-head), and T stands for *Tandem* (tail-head), for a total of nine possible orientation subgroups (AA, AC, DA, AT, TA, TT, TC, DT, an DC). For instance, the AC subgroup corresponds to *Reference features* where the upstream annotation is on *Any* strand, and the downstream annotation is on the opposite strand (*Convergent* orientation or in other words tail to tail, therefore the union of the subgroup TC and DC). Since all the *Reference features* are represented from left to right (5′–3′) in an aggregate (and individual) profile, those standing on the negative strand are virtually flipped (along with their flanking annotations) (*see* **Note 1**) such that it is impossible to get an upstream orientation *Convergent*.
26. Generating the heatmap images could take much longer than generating the actual aggregate and individual data files. Also, note that this process could use a lot of memory if you are working with a lot of *Reference features* and/or a lot of windows;

consequently, it is recommended to use the 64 bits version of Java to generate heatmap images.

27. In the exploratory analysis phase, it is preferable to keep an automatic scaling to generate the aggregate graphs, but for heatmap representation, it is preferable to set values for the Y-axis because of the outliers.
28. By pressing "Run," the interface creates the file "VAP_parameters.txt," copies the appropriate platform-dependent binary file (*see* **Note 2**) from the jar file in the selected output folder, and executes it from there. Any relative path must therefore take the output folder as the working directory when using the interface. Note that if an executable file named "vap_native" is present in the output folder, the interface rather executes this file. This "vap_native" strategy allows a user to compile the source code (released under the GPL license, https://bitbucket.org/labjacquespe/vap_core/) on a server with an unsupported OS (or any other computer) and use the interface through an X windowing system without having to repackage the jar file.
29. The name of the aggregate graph depends on the options (referred at Subheading 3.2 **idem steps 3e–h**) allowing to generate one graph per element (value of 1 in the next table) or to put all the elements on one graph (value of 0):

Dts	Grp	Ori	GraphFileName	DataFileName
1	1	1	agg_grph_dts_grp_ori	agg_dt_dts_grp_ori
1	1	0	agg_grph_dts_grp	agg_dt_dts_grp_ori
1	0	1	agg_grph_dts_ori	agg_dt_dts_ori
1	0	0	agg_grph_dts	agg_dt_dts_ori
0	1	1	agg_grph_grp_ori	agg_dt_dts_grp_ori
0	1	0	agg_grph_grp	agg_dt_dts_grp_ori
0	0	1	agg_grph_ori	agg_dt_dts_ori
0	0	0	agg_grph_all	agg_dt_dts_ori

Because the *Datasets* are always processed one at a time, in the case where you want all the *Datasets* on the same graph, the name of the "agg_data" files always contains the name of one *Dataset*, but the name of the graph does not because the different files are grouped through the list of aggregate graph to generate (idem for the orientation subgroup(s)) (*see* **Note 31**).

So far, one heatmap image and individual data file is generated for each combination of *Dataset* and *Reference group*; the

orientation of the flanking annotations is included as a column in the "ind_data" file (in the *Annotations* mode) but not used to create subgroups. The naming scheme is therefore equivalent to the 1-1-0 state but using different sub-prefixes.

30. As illustrated in Fig. 1a, b, the bottom graph shows that the proportion is ~45 % in the upstream and downstream regions, and it is dropping to ~20 % in the middle of the *Reference features*, while in Fig. 1c, the proportion is 100 % in the flanking regions and dropping to ~40 % in the middle. This indicates that in each window of the flanking regions (from panels 1a, b), ~45 % of the genes contribute a datavalue to the corresponding window of the upper graph (therefore ~55 % of missing data, mainly because of the sequencing depth that was not that high), while the parameter to process the missing data as datavalues of zeros in panel 1c implies to bring the proportion of contributing genes to 100 %, allowing a more reliable comparison of the *Datasets*. Note that, independently, by increasing the size of the windows, the proportion of genes contributing a signal datavalue automatically increases. The drop of the proportion in the middle of the graphs of Fig. 1 is the consequence of the genes being all split in the middle and each half aligned on one boundary (start or end) (*see* **Notes 19** and **20** for more details on the split). You can observe similar profiles by using the test *Datasets* provided with the interface (*see* Subheading 2.1 **item 3**).
31. The "agg_data" file mainly contains one line per window and four columns per *Reference group*: (1) the relative coordinates (where the 0 corresponds to the start of the *Reference feature* and the other coordinates are expressed in relative position to this *Reference point*, in steps corresponding to the size of the windows), (2) the aggregate value (*see* **Note 21**) after the smoothing (if applicable, *see* **Note 23**), (3) the dispersion value (*see* **Note 22**), and (4) the proportion of the *Reference features* of this group contributing to the aggregate value of this window (*see* **Note 30**). For example, in a case where a *Reference group* contains 500 *Reference features*, the aggregate value of a particular window is calculated on these (up to) 500 values; in the *Absolute* mode, if 100 *Reference features* do not have any data point in a particular window (e.g., shorter genes than the block capacity), the aggregate value is calculated on the 400 remaining values, and the column "proportion" contains this information (0.8 = 400/500). Note that the proportion information is useful to identify the regions of the aggregate profiles where the proportion of *Reference features* contributing to the aggregate value is too low to be reliable (e.g., if the capacity of a block is much bigger than the average gene length, the aggregate profile around the split point will be noisy and the propor-

tions very low because really few genes will contribute datavalues to these windows).

The "list_agg_graph" file contains the instructions for the Java interface on how to assemble the "agg_data" files to create the desired graphs. Briefly, the first column of this file contains the name of the graph, and the second column contains the name of the aggregate data file to include in this graph (based on Subheading 3.2 **idem steps 3e–h**).

32. It is important to notice that even the best data normalization cannot account for experimental variability such as antibody affinity directly affecting the ratio signal/noise. As a consequence, the dispersion value of the aggregate mean can be useful to compare the signal intensity of one *Dataset* over different *Reference groups* (e.g., gene expression groups) but not to compare different *Datasets* (e.g., different antibodies) over one *Reference group*. Comparing the same strain (or cell line) and antibody in different conditions (e.g., before and after treatment) could be more reliable, but in cases where there is an important biological impact (e.g., global depletion or amplification), the normalization could not work properly; in such case, the use of external spikes-in is recommended [20, 21].
33. The "ind_data" file mainly contains one line per *Reference feature* of a group, with the first nine columns including the *Reference feature* and region information (name, alias, coordinates, and orientation) followed by N columns containing the weighted average of the datavalue(s) of a given window (*see* **Note 18**), where N corresponds to the sum of all the windows per block requested by the user (*see* **Note 19**). As for the aggregate graphs, the 0 of the *X*-axis (represented in one of the *N* columns) corresponds to the start of the *Reference feature*, and the other coordinates are expressed in relative position to this *Reference point*, in steps corresponding to the size of the windows. At the bottom of the file, the aggregate values of the requested orientation(s) are also provided. Note that the values included in this file are not affected by the smoothing of the aggregate values.
34. This file contains all the required parameter preceded by the unique tag "~~@" and uses the structure "parameter_name=value" (e.g., "~~@analysis_mode=A" for the *Annotations* mode), as well as a small description including the accepted values for each parameter. The format of this file is based for now on the APT (Almost Plain Text) format defined by Apache Maven (http://maven.apache.org/doxia/references/apt-format.html), corresponding to a lightweight markup language easy to read by human and easy to convert into a common document markup language like HTML.

35. More precisely, the seven blocks correspond to (1) the upstream region, (2) the first exons, (3) the first introns, (4) all the middle exons, (5) the last introns, (6) the last exons, and (7) the downstream region. For genes having only one exon, the blocks three to six are empty, and with two exons, the only intron is considered as the first (not the last) and is therefore associated to the third block.
36. By using one window of a defined size (e.g., 1 kb) aligned to the left, you will get the average signal over this window size (up to the size of the feature) as described in Subheading 3.5 **step 3**. Alternatively, by using two (or more) windows in the second block and by using some of the split alignment parameters, it is also possible to calculate the average signal over certain portion of the genes (such as the first/last half/quarter or the first/last *X* bp of the features). You can also choose the *Relative* analysis method if you are only interested in the proportion of the feature (one window automatically represents 100 %, two equals to 50 %, etc.).
37. You can execute the vap_core module from your current directory by simply typing to the terminal `./vap_core -p param_file` where `param_file` is defined at Subheading 3.2 **step 5c**.
38. By using a list file, vap_interface does not have access to some required information that are available in the parameters file to add all the elements to (including the number of *Reference points* and the number of windows per block) such that the annotations under the X-axis are missing in the created graphs. Also note that when you use the parameters file, the value of any parameter defined on the interface is applied no matter the value present in the file.
39. The command is `java -jar vap-1.1.0-all.jar create_graphs [options]` where the options allow doing the same operations from the terminal than from the interface.

References

1. Coulombe C, Poitras C, Nordell-Markovits A et al (2014) VAP: a versatile aggregate profiler for efficient genome-wide data representation and discovery. Nucleic Acids Res 42:W485–W493
2. Zentner GE, Henikoff S (2014) High-resolution digital profiling of the epigenome. Nat Rev Genet 15:814–827
3. Kent WJ, Sugnet CW, Furey TS et al (2002) The human genome browser at UCSC. Genome Res 12:996–1006
4. Robinson J, Thorvaldsdóttir H, Winckler W et al (2011) Integrative genomics viewer. Nat Biotechnol 29:24–26
5. Meyer CA, Liu XS (2014) Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. Nat Rev Genet 15:709–721
6. Nordell MA, Joly BC, Toupin D et al (2013) NGS++: a library for rapid prototyping of epigenomics software tools. Bioinformatics 29: 1893–1894
7. Pokholok D, Harbison C, Levine S et al (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 122:517–527
8. Guillemette B, Bataille AR, Gévry N et al (2005) Variant histone H2A.Z is globally localized to the promoters of inactive yeast genes and regulates nucleosome positioning. PLoS Biol 3, e384

9. Holstege FC, Jennings EG, Wyrick JJ et al (1998) Dissecting the regulatory circuitry of a eukaryotic genome. Cell 95:717–728
10. The Encode Project Consortium, Dunham I, Kundaje A et al (2012) An integrated encyclopedia of DNA elements in the human genome. Nature 489:57–74
11. Hoffman MM, Ernst J, Wilder SP et al (2013) Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res 41:827–841
12. Bolger A, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120
13. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760
14. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359
15. Li H, Handsaker B, Wysoker A et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079
16. Lassmann T, Hayashizaki Y, Daub CO (2011) SAMStat: monitoring biases in next generation sequencing data. Bioinformatics 27: 130–131
17. Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842
18. Giardine B, Riemer C, Hardison RC et al (2005) Galaxy: a platform for interactive large-scale genome analysis. Genome Res 15: 1451–1455
19. Jeronimo C, Bataille AR, Robert F (2013) The writers, readers, and functions of the RNA polymerase II C-terminal domain code. Chem Rev 113:8491–8522
20. Bonhoure N, Bounova G, Bernasconi D et al (2014) Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. Genome Res 24:1157–1168
21. Lovén J, Orlando D, Sigova A et al (2012) Revisiting global gene expression analysis. Cell 151:476–482

Chapter 19

Circular Dichroism for the Analysis of Protein–DNA Interactions

Garry Scarlett, Giuliano Siligardi, and Geoffrey G. Kneale

Abstract

The aim of this chapter is to provide information on the practical aspects of circular dichroism (CD) and synchrotron radiation circular dichroism (SRCD) in protein–nucleic acids interaction solution studies. The chapter will describe the guidelines appropriate to designing experiments and conducting correct data interpretation, the use of both benchtop and synchrotron CD approaches is discussed and the advantages of SRCD outlined. Further information and a good general review of the field a can be found in Gray (Circular Dichroism of protein–nucleic acid interactions. In: Fasman GD (ed) Circular dichroism and the conformational analysis of biomolecules. Plenum Press, New York. pp 469–500, 1996).

Key words Circular dichroism, Synchrotron CD, DNA–protein interactions

1 Introduction

CD is the differential absorption of left and right circularly polarized light by chiral molecules, i.e., molecules that are non-superimposable on their mirror image (enantiomeric) forms. Modern CD spectrometers directly measure this differential absorption, where $\Delta\varepsilon$ is the resulting extinction coefficient whose units are M^{-1} cm^{-1} (*see* **Note 1**). Viewed alternatively, the differential absorption of left and right polarized light will cause the resultant beam to be elliptically polarized away from the incident plane and the angle between the two acts as a measure of the difference in absorption. Because of this, CD spectra are often reported in molar ellipticity $[\theta]$ expressed in millidegrees. In biological systems the asymmetric carbon atoms present in the sugars of nucleotides and in natural amino acids, with the exclusion of glycine, result in both nucleic acids and proteins displaying a CD signal. Further contributions to the optical activity of the polymers result from their ability to form well-defined secondary structures - in particular helices, which themselves possess asymmetry. As a consequence, CD has found

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_19,

widespread use in the estimation of protein secondary structure [1], using various algorithms of which CONTINLL [2], SELCON3 [3] and CDSSTR [4] are the most frequently used and available online.

Similar studies, though less widespread, have sought to correlate structural parameters of DNA with their CD spectra [5–7], with some success particularly in assigning quaternary structures to nucleic acids, for instance in the case of triplex DNA and G-quadruplex mediated structures [8, 9]. The relatively limited success of assigning structural parameters to nucleic acid conformation compared to protein secondary structure determination is largely a consequence of the presence of both purine and pyrimidine chromophores. The distinct purine (adenine and guanine) and pyrimidine (thymine, uracil and cytosine) chromophores result in nucleic acids with the same structure but different base composition displaying CD spectral differences, precluding the ease of secondary structure estimation exploited in proteins. However, the technique can be used qualitatively to assess the types of DNA structure observed in solution as a function of solvent composition, ionic strength, ligand binding and temperature. It therefore follows that the perturbation of nucleic acid and protein secondary structure by their interaction can be usefully monitored in solution by CD spectroscopy.

The overlap of the UV absorption bands of nucleic acids and proteins means that CD studies of protein–DNA interactions can be complicated by the contributions observed from both components. This is particularly true for wavelengths less than 250 nm. However, in practice the CD signal between 250 and 300 nm is due only to the nucleic acid, as the contribution arising from the protein aromatic side-chain chromophores at standard experimental concentrations is about two orders of magnitude lower than that required for detectable CD spectra. Therefore the CD bands in this range can be used as diagnostic of DNA conformational changes due to protein binding interactions. CD has been used for determination of the stoichiometry and the dissociation constants (K_d) of DNA–protein, ligand–protein, and protein–protein binding interactions [10–12]. However, for very high affinity interactions ($K_d \sim$ nM), it may not be possible to accurately measure the CD signal at sub-nanomolar concentrations, when other techniques such as fluorescence spectroscopy or EMSA may be preferred. As is the case for many biophysical techniques, the complexities of systems displaying multiple binding sites—which may have different affinities and are often cooperative—can also make the analysis problematic. Despite this in many cases CD remains a good option not only for determining the K_d and stoichiometry of protein–DNA interactions but also for providing qualitative information about the nature of the interaction (e.g., conformational changes on binding) [12].

In the case of the fd gene 5 protein, CD was used to show two distinct binding modes [13]. CD has also been used to show that conformational changes induced by the bound *Lac* repressor are different for operator DNA and for random sequence DNA [14]. Similar studies on the *Gal* repressor demonstrated the involvement of the central G-C base pairs of the operator sequence in repressor induced conformational changes [15]. Studies on the interaction of *cro* protein of bacteriophage λ have also revealed different conformational changes for specific and nonspecific DNA binding [16]. Despite the apparent lack of any direct interaction of the central base pair of the operator sequence with *cro* protein, base substitution at this site was shown to affect the CD spectrum considerably.

Some additional examples in which CD has been used to examine protein–DNA interactions include the SRY-related protein, sox5, and the controller protein, C.AhdI [17, 18]. Both these proteins cause the DNA to bend. Sox-5 is a DNA sequence specific protein that has a single globular domain which opens up the DNA by intercalation of the methionine side chain in helix 1. The N and C-terminal residues stabilize the bending [17] and as a consequence there is a large enhancement of CD intensity as well as wavelength maxima shift of the CD bands. On the other hand, C.AhdI is a dimer in free solution at CD concentrations; two dimers bind on to the operator sequence 35 bp operator sequence [18] causing a large enhancement of CD intensity without any wavelength maxima shift (*see* Fig. 1a). Hence, although both proteins promote DNA bending, the two proteins must bind in different ways.

It is interesting to note that the 35 bp operator to which C.Ahd binds has two sites, one for each dimer; binding to the left operator, O_L, is much stronger than that to the right operator O_R so this site is filled first. However, binding of a second dimer to O_R is highly cooperative and under most experimental conditions, only the tetrameric complex is seen. Nevertheless, it was possible to observe binding to the left hand operator, O_L, when O_R was mutated to a nonspecific sequence, which greatly decreases the affinity at this site [18]. Titrating the protein into the DNA and observing the change in CD signal (Fig. 1) shows that the binding of each dimer causes an identical change of CD intensity (i.e., the CD change is additive). This suggests that the structural change is localized to the two sites and that there is no additional perturbation to the DNA structure (at least none observable by CD) arising from the interaction between the two bound C.AhdI dimers.

A large (~50 %) increase in $\Delta\varepsilon$ arising from the DNA was also observed when a type I DNA methyltransferase bound to its DNA recognition sequence, indicating a large deformation of the DNA in the DNA–protein complex [18] (*see* Fig. 2). CD has, in addition,

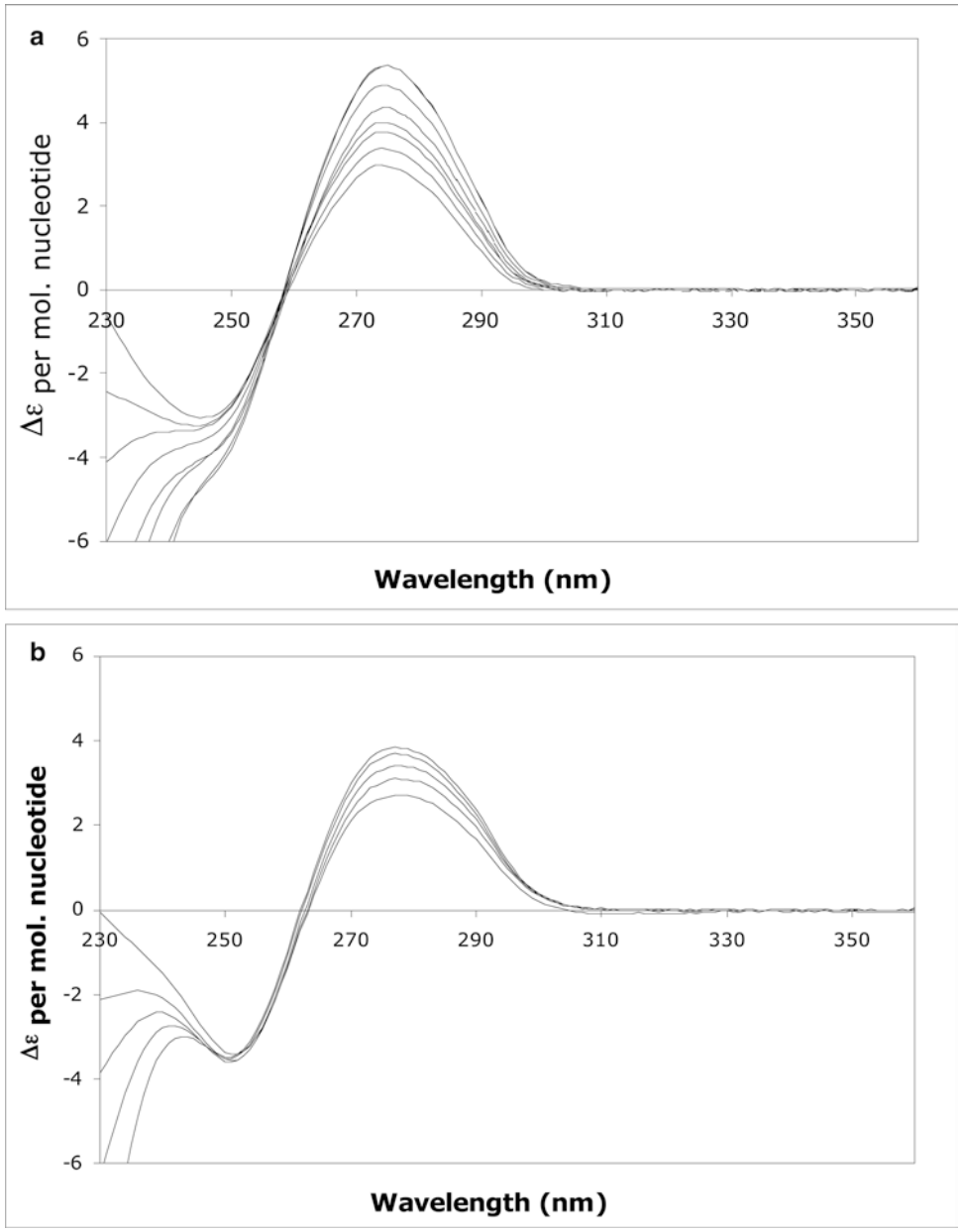


Fig. 1 CD spectra of the titration of protein C.AhdI into 35 bp oligonucleotide. (**a**) The 35 bp native operator sequence and (**b**) the 35 bp sequence in which the right operator has been mutated [16]. Four consecutive scans were measured for each addition of 50 μM aliquots of C.AhdI and averaged from 360 to 230 nm with data collected every 1 nm. Protein–DNA ratios of (**a**) 0, 0.54, 1.1, 1.6, 2.2, 3.0, 4.1, and 4.6, and (**b**) 0, 0.49, 1.0, 1.74, and 2.23 respectively. All CD spectra were corrected for dilutions due to ligand aliquot additions. Both titrations were carried out in 40 mM citrate pH 5.6, 100 mM NaCl, 1 mM EDTA at 20 °C in a 4 mm path-length cell (119-004F QS)

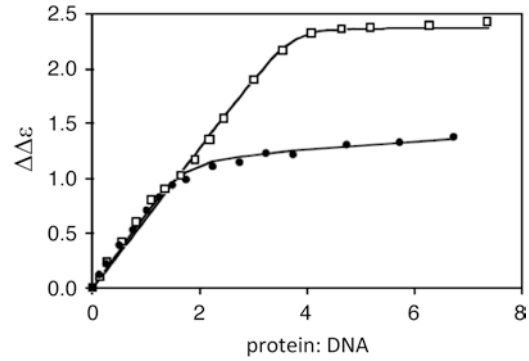


Fig. 2 Plots of the change in $\Delta\varepsilon$ (i.e., $\Delta\Delta\varepsilon$) at 275 nm (calculated by subtracting the corrected value of the DNA alone) versus the molar ratio of protein–native DNA (open squares) and protein–mutated DNA (*filled circles*) [18]. The values of $\Delta\varepsilon$ were taken at the peak maxima wavelengths (274.5 nm and 277.5 nm respectively) of the CD spectra shown in Fig. 1. Data were fitted to a two-step binding model with the binding affinity for the first site, $K_1 = 2 \times 10^7\ M^{-1}$ in both cases. For binding to the second site, the best fit was obtained with $K_2 = 3 \times 10^7\ M^{-1}$ for the native operator and $K_2 = 2 \times 10^4\ M^{-1}$ for the mutant operator, *see* [18] for further details

been shown to be a useful tool when studying RNA triplex structures [19], binding of peptides TAT and REV [20, 21] to RNA and the disordering of T7 RNA polymerase [22].

2 Materials

1. Good quality benchtop CD spectrometers such as the Applied Photophysics Chiroscan and Jasco J800 series spectrometers can detect to 190 nm when used for the study of protein and nucleic acids conformations in isolation (for setting up, *see* **Note 2**). However, for protein–nucleic acid binding interactions the far UV region is usually limited to 230 nm, which limits qualitative assessment of protein conformational changes induced by nucleic acid binding. A greater penetration into the far UV region, down to 200 nm, can be achieved using SRCD spectroscopy (*see* **Note 3**). These are unattainable with benchtop CD instruments without measuring spectral distortions. However, SRCD requires access to a synchrotron national central facility, such as that at the UK Diamond Light Source (http://www.diamond.ac.uk/Home.html).

2. A high quality quartz cell with low strain is required for accurate measurements. The cell path-length will depend on the absorption properties and concentration of the sample. The optimum UV absorption of a solution (including both solvent and solute) to give best signal-to-noise ratio is about 0.9 absorbance units. A wide range of cell pathlengths, such as 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, and 5 cm are available and enable measurements (following the Beer–Lambert Law to keep the absorbance below 0.9) for a broad range of sample concentrations. Path-lengths of 0.2–1.0 cm are usually recommended for measurement of the nucleic acid signal in the vicinity of 275 nm, where the signal is weak and buffer absorption is negligible. Protein secondary structure analysis requires measurements from 260 nm down to 180 nm (or lower) and the higher absorption in the lower end of this range typically require cell path-lengths of 0.1–2 mm, dependent on the sample concentration and the solvent absorption properties. However, the combined absorbance of the protein and DNA when studying their interaction at wavelengths below 230 nm can make CD readings inaccurate. Moreover, the spectra overlap considerably at this point and are difficult to deconvolute. Therefore interpretation of possible protein conformational changes induced by the DNA binding is difficult compared to monitoring conformational changes in the DNA induced by the protein. A better assessment of protein conformational changes resulting from the interaction can be achieved using SRCD beamlines, mainly because the CD measurements of protein–DNA complexes can be extended to much lower wavelengths in the far UV region than is be achievable with benchtop CD instruments (Fig. 3).
3. Buffers should be prepared using high quality reagents and water. Use buffers that have low absorbance in the wavelength region of interest. Tris–HCl, citrate, perchlorate, and phosphate are routinely used, normally at concentrations of 1–10 mM. As chloride anions absorb in the far UV region, phosphate buffers are preferred for measurements in this region and the use of fluoride anions such as NaF and KF should be used instead of NaCl. However, it is important to verify via other techniques, such as the gel retardation assay (EMSA—*see* Chapter 1), that changing the buffers in this way does not affect binding.
4. Stock solutions of the protein and nucleic acid should be in the same buffer, ideally achieved by dialysis or buffer exchange. The protein should be as concentrated as possible whilst avoiding aggregation so as to minimize dilution during the titration. If a synthetic DNA fragment containing the recognition sequence is to be used it should be close to the minimum size required for binding, to maximize the change in CD signal

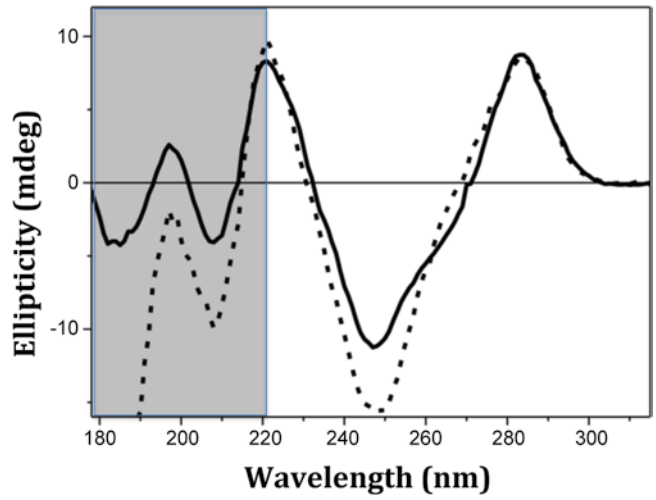


Fig. 3 CD spectra of a 21 bp oligonucleotide in 100 mM KF measured with a Chirascan Plus applied photophysics benchtop machine (*dashed line*) and SRCD B23 module B (*solid line*). Note the spectral distortions on the benchtop machine at high signal intensities that correlate with high HT voltages required across the detector. The *grey area* highlights the extended lower wavelength in the far UV region using SRCD that can be observed and is otherwise unattainable with benchtop CD instruments

and avoid nonspecific binding. Oligonucleotides should be purchased at least at the 200 nmol synthesis scale and preferably also HPLC purified. To avoid denaturation and degradation, keep concentrated solutions of protein and DNA frozen in small aliquots. Whilst working at the laboratory bench it is recommended to keep all solutions on ice to prevent aggregation or degradation and sealed to prevent condensation and evaporation.

5. A supply of dry nitrogen (oxygen free) is recommended for any CD measurement as this protects the optical components, prevents the production of ozone, reduces the HT voltage on the photomultiplier and improves the transmission below 220 nm. The supply may be from a gas bottle or boiled-off liquid nitrogen; both systems should have an absorption filter before the optical unit (*see* **Note 4**).
6. Calibration of a CD instrument in millidegrees ellipticity is achieved with a solution of (+)10-camphor-sulfonic acid at a concentration of 5.0 mg/ml. The concentration may be accurately checked by measuring the absorption in a UV spectrophotometer using a molar extinction coefficient of 34.5 at 285 nm.
7. Temperature control of the sample is essential when comparative measurements are being made or in experiments in which the instrument is used for long periods of time, for example a

titration or a kinetics experiment. Hence the usage of a Peltier cell with water cooling is an essential part of the spectrophotometer kit, maintaining the sample temperature ±0.1 ° C within the 4–90 °C.

3 Methods

For most proteins, there is no significant CD spectrum between 250 and 300 nm compared to that seen for nucleic acids. Experiments involving the addition of protein and observations of the DNA signal can thus be conveniently carried out in this wavelength range. Below 250 nm, the CD contributions of protein and DNA overlap; hence it may be necessary to compare the observed spectrum with that of the simulated spectrum, which can be calculated from the sum of the individual spectra of the protein and DNA. Deviation between the observed and calculated spectra should reveal evidence of a binding interaction.

A list of recommendations is outlined below to help the successful collection of good data devoid of artifacts and distortions.

1. To prevent damage to the optics, flush the instrument with nitrogen for 10–15 min before switching the lamp on. Continue to purge the instrument for the duration of the experiment.
2. Switch on the electronics first and then the lamp and allow the instrument to stabilize for 30 min. After warming-up is complete switch on the Peltier temperature controller and set the temperature to 20 °C allow a further wait of 10 min before making any measurements.
3. Whilst waiting for the instrument to warm up, measure the UV spectrum of both the DNA and protein (*see* **Note 5**); calculate the concentration of the stock solutions from their extinction coefficients. The stock solution of protein should be at as high a concentration as possible, to minimize corrections for dilution in subsequent titrations.
4. Measure the UV absorbance of the buffer in the cell (*see* **Note 6**) to ascertain the buffer has little or no absorption in the wavelength region in which the CD measurements are to be made.
5. Using the data from **steps 3** and **4**, make sure the total UV absorbance at the wavelength(s) of interest is not too high. Often the value of 0.9 Absorbance unit is taken as the upper limit for the protein–DNA titration end point but with modern CD instruments this can be extended to 1.5, if necessary. Absorption of >1.5 will lead to spectral distortions, and it is important to limit the absorption to below this value to obtain good quality data.

6. The choice of the cuvette pathlength to be used in the titration depends on the starting concentration of oligonucleotide, which in turn is dictated by the type of experiment to be undertaken. In general, for experiments to determine the stoichiometry of the interaction, it is a good idea to have a starting oligonucleotide concentration about 50 times the estimated dissociation constant, with the aim of achieving stoichiometric binding. However, to determine the dissociation constant of the interaction, the titration should be repeated at lower concentrations. The DNA concentration should ideally be at or below the K_d (which may have to be estimated by trial and error), in order to ensure that not all the DNA is bound at the stoichiometric point.
7. Clean the 0.1 cm cell (*see* **Note 7**) and calibrate the instrument using freshly prepared 1 mg/ml aqueous solution of (+) 10 camphorsulfonic acid (CSA) and scan the CD in the 185–350 nm wavelength range using a 1 nm bandwidth. The CD spectrum will show a positive band at 290.5 nm with ellipticity of 33.7 mdeg and a negative band at 192.5 nm with at least a magnitude of twice that at 290.5 nm. A well-calibrated CD instrument should have the ratio of the magnitude of the CD band at 192.5 over that at 290.5 to be between 1.9 and 2.1 ($1.9 > (192.5/290.5) < 2.1$); if not the instrument must be tuned appropriately as per the manufacturer instructions.
8. Take the cell set aside for the experiment, clean it and fill with buffer. Place the cell in the instrument taking care to note the orientation of cell in the beam. If using a cylindrical cell place it so that the neck of the cell rests against the side of the cell holder. Run a baseline between 200 and 350 nm, signal average over 9 scans to reduce the noise. This multiple scanning should show reproducible and superimposable CD spectra within the spectral noise and hence a constant condition has been achieved for recording the experiment.
9. Clean the cell and replace the buffer with the DNA solution and run the spectrum under the same conditions. When you remove the cell remember to place the cell back in the holder with the same face towards the light source as before and check the signal relative to that of the buffer in the same wavelength region 200–350 nm.
10. Pipette a small aliquot of the stock protein solution into the DNA solution in the cell and mix. Allow 5 min for equilibration before measuring the CD spectrum. The concentration of DNA to start the titration depends on whether the aim is to determine stoichiometry or K_d, higher DNA concentrations are required for determining stoichiometry but DNA concentrations should be kept low for estimating the K_d. In practice if the oligonucleotide is kept at 20–40 base-pairs, the stoichiometry of the

interaction will probably remain in the 1–2 binding site range. In all cases the DNA concentration together with that of the final protein concentration (at no less than a 3–4 molar excess) will dictate the choice of pathlength so as to obtain a maximum absorbance of no more than 1.5. The titration can be conducted either by adding small aliquots of a highly concentrated protein stock solution directly into the cuvette cell containing the DNA solution or preprepared individually mixed DNA–protein solutions each at increased protein molar ratios. In both cases it is recommended to prepare a protein stock solution that is 50 times more concentrated than that of the DNA in order to reduce the dilution factor upon addition of the protein. For initial experiments, the molar quantity of protein added at each step should be perhaps 10 % of that of the DNA. For every aliquot of protein added the base line must be checked for any variations and the problem must be resolved before the next addition is made (*see* **Notes 4** and **6**).

11. Repeat the addition of protein to the DNA until no further changes in the CD are observed.
12. Modern CD spectrometers are quite stable; however, for older instruments it is wise to check whether there is any baseline drifting by repeating the CD measurement of the buffer.
13. Select a spectral change and plot the measured CD parameter at a suitable wavelength against the concentration of protein added. When the titrations are conducted at concentrations above the K_d, the stoichiometry can be determined directly from the plot of CD intensity at fixed wavelength versus the protein–DNA ratio (or at sub-stoichiometric concentrations of DNA, by fitting to the appropriate binding equation).
14. Once the spectral changes and stoichiometry have been established, it is often useful to repeat the experiment with rather more titration points, using smaller aliquots of protein. Although the maximum CD signal from DNA is normally obtained around 275 nm, one should work at the wavelength that corresponds to the largest difference between free and bound DNA, which may well be different. The relevant equilibrium binding constants can be obtained by fitting the experimental titration curves to the appropriate binding isotherm [12].

4 Notes

1. $\Delta\varepsilon = \varepsilon_L - \varepsilon_R$ where ε is the molar extinction coefficient ($M^{-1}\ cm^{-1}$) for the left (L) and right (R) components. However, CD spectra are often expressed in ellipticity (θ). For determination (or comparison) of the secondary structure of polypeptides and polynucleotides, molarity is often expressed in terms of moles of amino acid or nucleotide residues respectively.

Molar ellipticity $[\theta]$ expressed in degrees M^{-1} cm^{-1} is related to $\Delta\varepsilon$ by the equation:

$$[\theta] = 3298\Delta\varepsilon\left(M^{-1}\,cm^{-1}\right).$$

2. A number of instrument parameters must be set up before any spectrum is scanned. These are (1) the entrance and exit slits that normally will be set between 1 and 2 nm for protein or DNA samples. The wider the slits are set, the more light enters the optical system and hence a better signal to noise ratio is achieved, but at the cost of poorer resolution. For instance, if the wavelength calibration of the instrument is set using the fine structure lines of the xenon lamp (460–490 nm region), the slits should be set to 0.1 nm as these line are extremely sharp and will only be resolved with a narrow setting. (2) The wavelength start and finish values. The wavelength range for protein, DNA, or RNA structure analysis is typically 340–200 nm but it will vary within this range dependent on the experiment and the solvent conditions. CD measurements at specific wavelengths are sometimes performed for kinetic or temperature melting experiments. However, it is the spectrum in its entirety that will reveal the full conformational behavior of the molecule. Therefore it is recommended to carry out even these studies scanning the whole wavelength range. (3) The number of accumulations required. This depends on how noisy the spectrum is, as multiple accumulations improve the signal to noise ratio. Typically a good baseline might require nine scans and a sample four scans. The improvement in signal–noise after repeated scans varies as the square root of the number of scans, e.g., four scans will reduce the noise by half but it will take four times longer to complete the measurement. (4) The data collection settings. Data collection is dependent on how the manufacturer has designed the instrument, with the pi*-180 or Chiroscan from Applied Photophysics, we recommend collection at 0.5 or 1 nm intervals with the number of samplings set to 10,500, operating in conjunction with adaptive sampling, set error to ±0.01 and a maximum sampling of 500,000. For the Jasco J700 and J800 series, set the step resolution to 0.2 nm (this allows a maximum wavelength range of 400 nm), and a scanning speed of 20 nm/min and a response time of 4 sec for J700 and 50 nm/min for. Software packages from both manufacturers allow for smoothing of data, baseline subtraction, zeroing offsets and provide simple mathematical options to convert millidegrees into ellipticity per mole residue, or from millidegrees to delta absorbance and into delta molar extinction coefficient. The data can then be exported to Microsoft Excel or other similar software packages for dilution corrections and analyzed with a graphical output. (5) The temperature is normally set at 20 °C for most experiments.

On measuring several spectra after baseline subtraction it is normal that an overall baseline drift (or offset) away from zero can be observed. In this case you may use the offset in the wavelength region devoid of any signal to return to baseline. Any difference in the position of the cell in the cell holder can result in an offset. Ensuring the cell is always placed to present the same section to the beam each time can minimize the effect. However, in this region no change of buffer baseline *slope* should be corrected as this indicates light scattering artifacts. Light scattering is usually a consequence of sample aggregation that can usually be reduced by modifications in buffer and concentration.

3. Diamond B23 beamline [http://www.diamond.ac.uk/Beamlines/Soft-Condensed-Matter/B23.html] for synchrotron radiation circular dichroism (SRCD) provides two main advantages over conventional benchtop CD instruments: higher photon flux and highly collimated microbeam light. The high photon-flux enables better measurements in the vacuum UV (VUV) and far-UV regions, especially for samples in physiological conditions of high salt concentration otherwise unattainable even with good benchtop CD instruments. For nucleic acids in high salt concentrations, SRCD spectra can better reveal conformational differences and allow a more accurate estimation of secondary structure content. The highly collimated and small focal spot (0.6 mm × 0.4 mm) at the sample enables CD measurements using small aperture cuvette cells with small volume capacity (20 μl for 1 cm cell to 780 μl for 10 cm cell) [23] which is important for precious samples difficult to produce in large quantities.
4. High intensity UV radiation converts oxygen to ozone, which damages the optics. Failure to purge will lead to deterioration in instrument performance and be detrimental to the health of those working in the vicinity of the instrument. Flushing the optics with nitrogen gas can be done from a liquid nitrogen Dewar (Taylor-Wharton XL-45) with the gas flow at 4 l/min; this gives 3 weeks of continuous purging and the instrument can then be purged for 3 days prior to measurements for the best results. Alternatively one can flush with a nitrogen (oxygen free) bottle, which will last up to 2 days before a replacement is required.
5. It is important to know the exact concentrations of DNA and protein to be used in the titration, these values can be determined by UV spectroscopy and the application of the Beer–Lambert law. For a precise calculation, it is best to experimentally determine the extinction coefficient of the DNA, as this can be quite different to the value calculated from the sum of the free bases, due to stacking interactions between the bases when

in a double helix (in fact there is also a smaller effect for single-stranded DNA). The extinction coefficient can be experimentally calculated by measuring the change in absorbance at 260 nm upon digestion of the DNA by phosphodiesterase. The percentage increase (i.e., the hyperchromicity) can then be applied as a correction factor to the calculated extinction coefficient. However, as stacking interactions in proteins are very rare, the extinction coefficient in this case can be taken simply as the sum of the extinction coefficients of the constituent aromatic residues.

6. For protein–DNA interaction studies by CD spectroscopy, it is recommended to scan the sample in the 340–200 nm wavelength region, although it may be difficult to keep the absorbance low enough to acquire good data in the sub 250 nm range. To reach the far UV limit of 200 nm, the amount of salt, Na and K chloride ions present in solution will dictate the choice of the pathlength cell to be used. Higher chloride ion concentrations will require a narrow pathlength of 0.05–0.1 cm, whereas lower salt concentrations might allow longer pathlengths of 0.5–1 cm pathlengths. The spectrum of the buffer used to prepare the DNA and protein solutions should always be scanned first to determine the right pathlength choice. Once the pathlength cell with a good transparent 330–200 nm region has been identified the concentration of the DNA and protein ligand can be calculated from their respective molar extinction coefficients to obtain a maximum UV absorbance of no more than 1.5 upon completion of the titration.
7. It is important to clean thoroughly the cell after each sample measurement, washing the cell copiously with highly deionized water followed by 96 % ethanol. Wipe the outer surface of the cell with soft tissue paper before blow-drying the inside of the cuvette cell with nitrogen gas. For cuvette cells of narrow pathlengths, the cleaning might be repeated several times and/or conducted using detergent base cleaning solutions (Hellmanex® II).

References

1. Gray DM (1996) Circular dichroism of protein-nucleic acid interactions. In: Fasman GD (ed) Circular dichroism and the conformational analysis of biomolecules. Plenum, New York, pp 469–500
2. Provencher SW, Glöckner J (1981) Estimation of globular protein secondary structure from circular dichroism. Biochemistry 20:33–37
3. Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem 287: 252–260
4. Van Stokkum IHM, Spoelder HJW, Bloemendal M, Van Grondelle R, Groen FCA (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. Anal Biochem 191:110–118
5. Johnson BB, Dakl KS, Tinoco I Jr, Ivanov VI, Zhurkin VB (1981) Correlations between Deoxyribonucleic Acid Structural Parameters and Calculated Circular Dichroism Spectra. Biochemistry 20:73–78

6. Basham B, Schroth GP, Ho PS (1995) An A DNA triplecode: thermodynamic rules for predicting A and B DNA. Proc Natl Acad Sci U S A 92:6464–6468
7. Scarlett GP, Elgar SJ, Cary PD, Noble AM, Orford RL, Kneale GG, Guille MJ (2004) Intact RNA-binding domains are necessary for structure-specific DNA binding and transcription control by CBTF^{122} during Xenopus development. J Biol Chem 279(50):52447–52455
8. Gray DM, Hung SH, Johnson KH (1995) Absorption and circular dichroism spectroscopy of nucleic acid duplexes and triplexes. Methods Enzymol 246:19–34
9. Hardin CC, Henderson E, Watson T, Prosser JK (1991) Monovalent cation induced structural transition in telomeric DNAs: G-DNA folding intermediates. Biochemistry 30:4460–4472
10. Siligardi G, Panaretou B, Meyer P, Singh S, Woolfson DN, Piper P, Pearl LH, Prodromou C (2002) Regulation of Hsp90 ATPase activity by the co-chaperone Cdc37p/p50(cdc97). J Biol Chem 277:20151–20159
11. Siligardi G, Hussain R (2010) Applications of circular dichroism in encyclopedia of spectroscopy and spectrometry, vol. 1, 2nd ed. In: Lindon J, Tranter G, Koppenaal D (eds). Elsevier, Oxford. pp 9–14
12. Carpenter ML, Kneale GG (1991) Circular dichroism and fluorescence analysis of the interaction of Pf1 gene 5 protein with poly (dT). J Mol Biol 27:681–689
13. Kansy JW, Cluck BA, Gray DM (1986) The binding of fd Gene 5 protein to polydeoxyribonucleotides: evidence from CD measurements for two binding modes. J Biomol Struct Dynam 3:1079–1110
14. Culard F, Maurizot JC (1981) Lac repressor-lac operator interaction. Circ Dichroism Study Nucleic Acids Res 9:5175–5184
15. Wartell RM, Adhya S (1988) DNA conformational change in Gal repressor-operator complex: involvement of central G-C base pair(s) of dyad symmetry. Nucleic Acids Res 16:11531–11541
16. Torigoe C, Kidokoro S, Takimoto M, Kyoyoku Y, Wada A (1991) Spectroscopic studies on lambda cro protein-DNA interactions. J Mol Biol 219:733–746
17. Conner F, Cary PD, Read C, Preston NS, Driscoll PC, Denny P et al (1994) DNA Binding and bending properties of the post-meiotically expressed Sry-related Protein Sox-5. Nucleic Acids Res 22: 3339–3346
18. Papapanagiotou I, Streeter SD, Cary PD, Kneale GG (2007) DNA structural deformations in the interaction of the controller protein C.Ahdl with its operator sequence. Nucleic Acids Res 35:2643–2650
19. Taylor IA, Davis KG, Watts D, Kneale GG (1994) DNA binding induces a major structural transition in a type I methyltransferase. EMBO J 13:5772–5778
20. Pinhero P, Scarlett GP, Rodger A, Rodger PM, Murray A, Brown T, Newbury S, McClellan JA (2002) Structures of CUG repeats in RNA. J Biol Chem 277:35183–35190
21. Calnan BJ, Biancalana S, Hudson D, Frankel AD (1991) Analysis of the arginine-rich peptides from the HIV TAT protein reveals unusual features of RNA-protein recognition. Genes Dev 51:201–210
22. Tan R, Frankel AD (1995) Structural variety of arginine-rich RNA-binding peptides. Proc Natl Acad Sci U S A 92:5282–5286
23. Hussain R, Jávorfi T, Siligardi G (2012) Spectroscopic analysis: synchrotron radiation circular dichroism in comprehensive chirality, vol. 8. In: Carreira EM, Yamampto H (eds). Elsevier, Amsterdam. pp 438–448.

Chapter 20

Quantitative Investigation of Protein–Nucleic Acid Interactions by Biosensor Surface Plasmon Resonance

Shuo Wang, Gregory M.K. Poon, and W. David Wilson

Abstract

Biosensor-surface plasmon resonance (SPR) technology has emerged as a powerful label-free approach for the study of nucleic acid interactions in real time. The method provides simultaneous equilibrium and kinetic characterization for biomolecular interactions with low sample requirements and without the need for external probes. A detailed and practical guide for protein–DNA interaction analyses using biosensor-SPR methods is presented. Details of SPR technology and basic fundamentals are described with recommendations on the preparation of the SPR instrument, sensor chips and samples, experimental design, quantitative and qualitative data analyses and presentation. A specific example of the interaction of a transcription factor with DNA is provided with results evaluated by both kinetic and steady-state SPR methods.

Key words Biosensor, Surface plasmon resonance, Protein–nucleic acid interaction, Kinetics, Steady state analysis, Mass transfer, Transcription factor

1 Introduction

1.1 Surface Plasmon Resonance

During the past 20 years, commercial biosensors using surface plasmon resonance (SPR) detection have been introduced to the scientific community and have emerged as a major and powerful approach for characterizing biomolecular interactions with high quality kinetic and thermodynamic information [1–4]. From the initial development on protein–protein interactions, applications of SPR have been significantly extended to a diverse range of biomolecular complexes, including those involving protein–nucleic acid, protein–small molecules, and nucleic acid–small molecules [5–8]. In the SPR method one component of an interaction is immobilized on a sensor chip to create the biosensor interaction surface. The other component(s) of the interaction is (are) then injected over the sensor surface in a solution at the desired ionic strength and pH. Upon complex formation on the sensor surface between these species, the refractive index changes are converted

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_20,

into SPR responses. The real time responses allow the extent and rates of complex formation to be quantitatively determined. SPR detection offers a number of advantages over other bioanalytical technologies, such as optical methods for systems involving strong interactions and/or low fluorescence and absorbance. In addition, SPR generally requires only picomole to nanomole quantities of material which is minimal compared to other techniques used for the evaluation of biomolecular interactions.

Despite the widespread use of SPR in characterizing protein–ligand and protein–protein interactions, SPR analysis of protein–DNA complexes has been less extensive. Protein–DNA interactions present specific challenges for SPR characterization due to their generally high binding affinity and strong electrostatic nature. These features give rise to several practical complications: (1) mass transfer limits on kinetics, where the rate of transfer of components from the injected solution to the immobilized component is slower than the association reaction, (2) very slow dissociation coupled with rebinding during the dissociation phase, and (3) limited time for the association reaction due to volume limitations in the injection syringe. These limitations have initially stalled the adoption of SPR for the experimental study of protein–DNA interactions. More recently, several approaches have been devised to mitigate these challenges, including the use of high flow rates, low immobilization densities, and the addition of DNA in the flow solution [9–12].

To illustrate the utility of SPR in protein–DNA interactions, and methods for optimizing experimental conditions for challenging systems, the interaction between the transcription factor PU.1 (Spi-1) with a specific DNA sequence will be used as an example in this protocol. PU.1 is a member of the ETS-family of proteins which comprise an evolutionarily conserved family of transcription factors [13]. ETS-family proteins regulate the expression of a functionally diverse array of genes throughout the Metazoan kingdom [14–16]. Aberrant expression of ETS-regulated genes is implicated in many human and veterinary cancers [17–20]. All ETS proteins share a structurally conserved DNA binding domain (known as the ETS domain) that recognizes DNA sequences containing a central 5′-GGAA/T-3′ consensus [21]. ETS-DNA complexes are universally characterized by the insertion of an essential recognition helix in the major groove at the core consensus, while the flanking bases are recognized via backbone contacts [22]. Since ETS factors share only limited interchangeability in vivo [23–25], the mechanism by which they achieve biological specificity continues to be an active area of inquiry. Factor-specific interactions with their target DNA represent major specificity determinants, as evidenced by the strong correspondence of sequence preference of ETS domains in vitro to genomic occupancy of native ETS proteins in vivo [21, 26]. In the case of PU.1, recent thermodynamic and kinetic

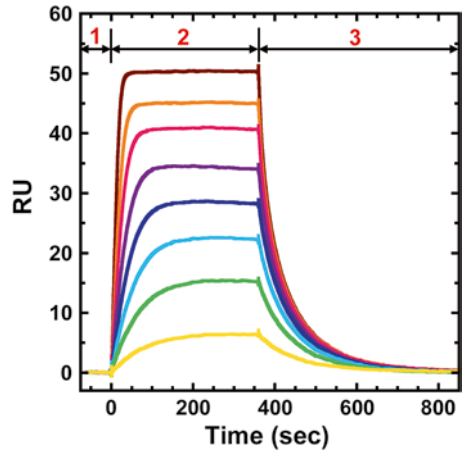


Fig. 1 Representative SPR sensorgrams with three major stages: (1) initial buffer flow for a stable reference baseline; (2) ligand association phase with sample injections over the sensor chip surface; (3) ligand dissociation phase with buffer flow over the surface

studies of DNA site recognition have shed light on the mechanism of DNA recognition [27–29], and established a molecular paradigm against which other ETS proteins may be quantitatively compared. Sequence-specific PU.1 ETS-DNA interactions, thus, represent an excellent model system for evaluating protein–DNA complex formation by SPR in terms of both equilibrium and kinetic signatures. The techniques described for ETS–DNA interactions should be highly transferrable to the investigation of other sequence-specific protein–DNA and peptide–DNA interactions.

1.2 Basic Principles for Biosensor-SPR Method

The results of a biosensor-SPR experiment are typically presented as a series of sensorgrams, which show the SPR response units (RU) as a function of time (Fig. 1). With a DNA sequence immobilized on the chip surface, initial buffer flow gives rise to a reference baseline. Subsequently, a protein solution is injected and as the solution flows over the surface, protein binding to DNA is monitored by changes in SPR response. With sufficient time a steady-state plateau is established where association and dissociation of protein achieves a steady state. Finally, buffer flow (without protein) is restarted and the dissociation of the complex can be monitored as a function of time (Fig. 1).

For a protein (P) binding to a DNA sequence and forming a single complex (C), the interaction is given by:

$$\mathrm{P} + \mathrm{DNA} \underset{k_{\mathrm{d}}}{\overset{k_{\mathrm{a}}}{\rightleftharpoons}} \mathrm{C} \qquad (1)$$

and the equilibrium binding affinity for this interaction is:

$$K_A = \frac{[C]}{[P][DNA]} = \frac{k_a}{k_d} = \frac{1}{K_D} \quad (2)$$

where [P] is the concentration of the unbound protein at the sensor surface, [DNA] is the concentration of the immobilized DNA which is not bound to protein (i.e., free DNA concentration), and [C] is the concentration of the protein/DNA complex; K_A is the equilibrium binding constant, k_a is the association rate constant, and k_d is the dissociation rate constant.

For association:

$$\frac{d[C]}{dt} = k_a [P][DNA] \quad (3)$$

and for dissociation:

$$-\frac{d[C]}{dt} = k_d [C] \quad (4)$$

Both the association and dissociation phases of the sensorgram can be simultaneously fitted from several sensorgrams at different protein concentrations using a global fitting routine [30, 31]. Global fitting allows the most robust determination of the kinetic constants and the calculation of equilibrium constant, K_A, from the ratio of kinetic constants (Eq. 2). Alternatively, steady state binding results from SPR experiments, where the SPR response reaches a plateau region, can be fitted with the following model:

$$r = \frac{nK_A C_{free}}{1 + K_A C_{free}} = \frac{RU_{obs}}{RU_{max}}$$
$$\left(\text{limit of } r \text{ as } C_{free} \to +\infty,\ r \to = n\right) \quad (5)$$

where n is the stoichiometry and is one for the single-site model, r represents the moles of bound protein per mole of DNA, and C_{free} is the free protein concentration at equilibrium with the complex. When C_{free} is very high, r approaches n. RU_{obs} is the observed (experimental) response in the plateau region and RU_{max} is the predicted maximum response for a monomer protein binding to a DNA site [8]. RU_{max} can be calculated and determined experimentally at the RU for saturation of the DNA binding sites, or used in Eq. 5 as a fitting parameter such that K_A, r, and RU_{max} are determined by fitting RU_{obs} versus C_{free}. Dividing the observed steady-state response RU_{obs} by RU_{max} at saturation yields the binding stoichiometry.

If the complex dissociates slowly, the surface can be regenerated before the complete dissociation occurs with a solution that causes rapid dissociation of the complex without irreversible damage to the immobilized DNA [31, 32]. For example, a solution at low or high

pH (pH ≤ 2.5 or pH ≥ 10) can unfold DNA and cause the complex to completely dissociate. Additional injections of buffer at pH near 7 allow the immobilized DNA to refold for additional steps in the binding experiment. After the dissociation and regeneration phase, a stable baseline is reestablished. Then another sample containing a different concentration of protein can be injected to generate another sensorgram. With a series of sensorgrams generated over a broad range of concentrations, both the kinetics and equilibrium constant can be determined.

1.3 Critical Factors for Protein–DNA Interaction Evaluation by Biosensor-SPR Methods

1.3.1 Concentration Range and Binding Affinity K_D

For the determination of an equilibrium constant by any method, the selected set of experimental concentrations must provide both free and bound concentrations of reactants. In the biosensor-SPR method, with DNA bound to the surface, the protein concentrations should span from below to above the K_D so that a range of bound states of DNA is obtained. The initial concentrations have little bound protein but as the concentration of protein injected is increased, the fraction of DNA saturation approaches one. In this way the most accurate equilibrium constant can be obtained. The sensorgrams will go from a very low RU and will approach a saturation value at high protein concentration relative to K_D. If too low a set of protein concentrations is used, all sensorgrams show sub-saturating levels of binding and it will be hard to determine the RU_{max}. If too high a set of concentrations is used, all sensorgrams will be near the saturation limit and global fitting may fail. Some preliminary experiments may be required with any new protein–DNA binding system to get an approximate K_D so that an appropriate set of protein concentrations can be prepared.

1.3.2 Mass Transport and Rebinding in Dissociation

For interactions on a sensor surface, the reaction component in sample solution is injected over the flow cell surface and must be transported from the bulk solution to the surface, a condition known as mass transport. In the reaction of interest, a protein, such as PU.1, is transported to the immobilized DNA for complex formation. This is a diffusion-controlled process, and the transport rate can directly influence the apparent binding kinetics if it is significantly slower than the binding reaction. A key requirement in accurate determination of kinetic constants by the SPR method is that the concentration of free protein in the matrix rapidly equilibrates with the flow solution. If the association reaction is much faster than mass transport, the observed binding will be limited by the mass transport processes. Conversely, if transport is fast and association is slow, the observed binding will represent the true interaction kinetics [33]. Therefore, the mass transport rate is a critical factor that must be considered in biosensor experimental design and in evaluating kinetic constants from biosensor-SPR methods.

A similar factor that affects kinetics is protein rebinding as a result of slow dissociation. After the protein–DNA complex

formation reaches the injection time limit, the experimental buffer flows over the sensor surface to dissociate the bound protein from the immobilized DNA. For proteins with very high binding affinities, they may rebind DNA after dissociation before leaving the matrix. When rebinding occurs the rate of protein transport away from the surface is slow compared to the real dissociation rate at the surface, and results in a very slow apparent dissociation. He et al. presented a theoretical and experimental approach to deal with the mass transfer effect on strong binding of proteins to DNA [9]. They used the lac repressor–operator interaction as a test system and compared their results with the previous SPR data that did not agree with filter binding results [10, 11]. Their data showed that, with the target DNA immobilized and lac repressor protein in the injection solution, strong mass transport effects and rebinding in dissociation were observed at low flow rates and high density of DNA immobilization. As the flow rate increased, the observed association rate constants were significantly increased in agreement with the elimination of mass transport effects. Similarly, in the dissociation phase of the sensorgram, rebinding of the dissociated protein markedly restricted the determination of the true dissociate rate. The authors designed a clever method to deal with the rebinding that includes excess lac DNA in the flowing buffer for dissociation. Using this method, the dissociated lac repressor protein bound to the DNA sequence in the flowing buffer, instead of the immobilized DNA on the surface, and consequently the rebinding was relieved and the real dissociation rate was determined [9].

Overall, for kinetic measurements, it is generally recommended to use low surface densities of the immobilized DNA and high protein flow rate (≥50 μL/min) to minimize the limitations on binding rates by mass transport processes. In addition, the dissociation phase can be set up for several hours or even longer with Biacore SPR which allows at least 50 % of bound protein to dissociate and a reliable kinetic fit can be performed, even with very slow dissociation.

1.3.3 Limited Time for the Association Due to Volume Limitations in the Injection Syringe

The volume of the injection syringe of a biosensor SPR system is limited, and thus the sample injection and association time is consequently limited. For example, the Biacore T200 instrument has an injection volume of 350 μL (syringe size), which obviously means that the association time can only be set to 7 min maximum with a flow rate of 50 μL/min, or 3.5 min with a flow rate of 100 μL/min. These limited association times are not sufficient for some interactions to reach the binding equilibrium necessary for steady state analyses. Myszka et al. showed that this problem can be resolved by immobilizing the target DNA sequence and placing the protein at different concentrations in the experimental buffer [9]. Then, protein samples are able to be injected for hours and the

kinetics can be determined with limited effects from mass transfer and rebinding. This method also allows the steady-state plateau to be reached and the K_A to be determined directly (Eq. 5) without any possible mass transfer issues.

In our biosensor-SPR evaluation of the interaction of PU.1 ETS domain with DNA, both mass transfer and rebinding are carefully evaluated and minimized in the experimental protocol detailed below. The incorporation of optimally designed flow cells in current instrumentation and optimized experimental protocols and sensor chips have qualified biosensor-SPR as an excellent method for quantitative analysis of protein–DNA interactions, especially for strong binding systems. Here, we show that careful use of ionic conditions allows useful data collection over a broad range of conditions without limiting mass transfer or rebinding problems.

2 Materials

2.1 Instrument Cleaning

These materials are for the Biacore T200 instrument but similar materials are required for other instruments.

1. Maintenance chip with a glass flow cell surface for cleaning without damage to experimental sensor chips.
2. 0.5 % (w/v) sodium dodecyl sulfate (SDS, Biacore Desorb solution 1).
3. 50 mM glycine pH 9.5 (Biacore Desorb solution 2) (*see* **Note 1**).
4. 1 % (v/v) acetic acid solution.
5. 0.2 M sodium bicarbonate solution.
6. 6 M guanidine HCl solution.
7. 10 mM HCl solution (*see* **Note 2**).

2.2 Sensor Chip Preparation Solutions for DNA Immobilization

1. A CM4 or CM5 sensor chip that has been equilibrated at room temperature for at least 30 min prior to use (all sensor chips are available from GE Healthcare Inc.) (*see* **Note 3**).
2. NHS: 100 mM *N*-hydroxysuccinimide freshly prepared in purified water.
3. EDC: 400 mM 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride freshly prepared in purified water.
4. Streptavidin immobilization buffer: 10 mM sodium acetate buffer pH 4.5.
5. 200 μg/mL streptavidin solution in streptavidin immobilization buffer.
6. 1 M ethanolamine hydrochloride in water pH 8.5.
7. Flow buffer during immobilization and chip activation (Running buffer) (HBS-EP): 10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.05 % (v/v) surfactant P20.

2.3 DNA Immobilization Solutions for a SA Chip

1. A streptavidin coated sensor chip (SA chip: prepared as described below or purchased) that has been equilibrated at room temperature for at least 30 min.
2. Running buffer (HBS–EP): 10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.05 % (v/v) surfactant P20 (as in Subheading 2.2).
3. Wash solution for prepared SA chip 1 M NaCl in 50 mM NaOH.
4. Biotin-labeled nucleic acid solution: 20 nM of 5′-biotinylated single strand or hairpin DNA dissolved in HBS-EP buffer.

2.4 Flow Solutions: General Buffers

1. HBS-EP (*see* **Note 4**): 10 mM HEPES pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.05 % (v/v) surfactant P20.
2. 10 mM Tris adjusted to pH 7.4 with HCl 100 mM NaCl, 1 mM EDTA, 0.05 % (v/v) surfactant P20.
3. 25 mM Na_2HPO_4 pH 7.4, 400 mM NaCl, 1 mM EDTA, 0.05 % (v/v) surfactant P20. Other NaCl concentrations can also be used.

2.5 Regeneration Solution

1. Generally used regeneration solutions are listed in Table 2. In general, milder conditions are initially used, and more harsh conditions are applied as needed. Some other regeneration solutions for special samples are available from the Biacore website. In our studies, 1 M NaCl solution is typically used as a gentle but efficient regeneration solution to remove protein from the DNA immobilized sensor chip surface (*see* **Note 5**).

3 Methods

3.1 Sensor Chip-SA Preparation for DNA Immobilization

1. Dock the CM4 or CM5 chip, then Prime with running buffer (HBS-EP). Start a sensorgram in all flow cells with a flow rate of 5 μL/min until baseline is stable (drifting < 1 response unit (RU)/min). "Dock" and "Prime" are Biacore control software commands that instruct the instrument to carry out specific operations. The commands and corresponding functions are listed in Table 1.
2. Mix 100 μL of NHS and 100 μL of EDC into one vial.
3. Inject the mixture of NHS/EDC for 10 min (50 μL) to activate the carboxymethyl surface to reactive esters.
4. Use Manual Inject to inject streptavidin solution over all flow cells with a flow rate of 5 μL/min for 20 min (100 μL). Track the RU immobilized, which is available in real time readout, and stop the injection after the desired level is reached (typically 2500–3000 RU for CM5 chip and 1000–1500 RU for CM4 chip).
5. Inject ethanolamine hydrochloride for 10 min (50 μL) to deactivate any remaining reactive esters.

Table 1
Biacore instrument commands[a]

Biacore control software commands	Function
Desorb	Removes adsorbed materials from the flow system
Sanitize	Removes disinfects from the flow system
Prime	Strongly flushes the flow system with running buffer
Dock	Docks the sensor chip into the instrument
Undock	Undocks the sensor chip from the instrument

[a]These commands are for Biacore instruments, but other commands with the same functions might be used with other instruments

6. Prime several times to ensure surface stability.
7. Then the sensor chip is ready for DNA immobilization as described under Subheading 3.2 (*see* **Note 6**).

3.2 DNA Immobilization on a SA chip

1. Dock a streptavidin-coated chip (SA chip) and start a sensorgram with a 25 μL/min flow rate.
2. Inject activation buffer for 1 min (25 μL) five to seven times to remove any unbound streptavidin from the sensor chip.
3. Prime several times to ensure surface stability.
4. Allow buffer to flow at least 5 min before immobilizing the nucleic acids.
5. Start a new sensorgram with a flow rate of 1 μL/min by choosing only one flow cell under "flow path" (e.g., flow cell 2 (FC2)) on which to immobilize the nucleic acid. Note not to immobilize nucleic acid on the flow cell chosen as the control flow cell. Generally, flow cell 1 (FC1) is used as a control and is left blank for subtraction, and different nucleic acids are immobilized on the remaining three flow cells (FC2-4).
6. Wait for the baseline to stabilize which usually takes a few minutes. Use Manual Inject, load the injection loop with ~100 μL of a 20 nM nucleic acid solution and inject over the current flow cell.
7. The amount of DNA to immobilize on the sensor chip depends on the relative molecular weights of the target DNA and protein and on the sensitivity of the biosensor system. Since the SPR response is directly proportional to the mass concentration of material at the surface, the theoretical protein binding capacity for a 1:1 interaction of a given surface is related to the amount of DNA immobilized [12]:

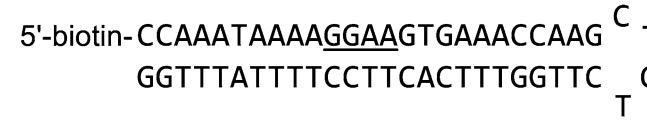


Fig. 2 5′-Biotin-labeled target hairpin DNA sequence for PU.1

$$\text{protein binding capacity}(\text{RU}) = \frac{\text{MW of protein}}{\text{MW of DNA}} \times \text{immobilized DNA level}(\text{RU}) \quad (6)$$

For example, the molecular weight of DNA in Fig. 2 is 16,962 Da and the molecular weight of the PU.1 ETS domain is 12,000 Da. Immobilizing 100 RU of DNA will give a theoretical PU.1 binding capacity of 70 RU assuming that the PU.1 is 100 % bound in a 1:1 complex (*see* **Note 7**).

8. Track the RU immobilized and stop the injection after the desired level is reached (typically ~300 RU for hairpin nucleic acid in ~20–30 bases in length, and ~100 RU for hairpin nucleic acid containing ~50–60 bases for kinetic experiments to minimize the mass transport effects).
9. At the end of the injection and after the baseline has stabilized, use the instrument's crosshair to determine the RUs of nucleic acid immobilized. The amount of nucleic acid immobilized is required to determine the theoretical moles of protein binding sites for the current flow cell.
10. Repeat **steps 5–9** for other flow cells (e.g., FC3 and FC4).
11. Stop the sensorgram and change the running buffer for the experimental buffer and Prime the system four times.
12. Before the first experiment, leave the sensorgram running at experimental buffer overnight at 10 μL/min in order to stabilize the baseline of the newly prepared sensor chip.

3.3 Sample Preparation

1. The sample solution must be prepared in the same buffer used to establish the baseline (the experimental buffer). If the protein requires the presence of a reducing agent, such as DTT (dithiothreitol) or TCEP (Tris(2-carboxyethyl)phosphine), to prevent oxidation of free cysteines, or nonspecific DNA (NS DNA) to reduce the nonspecific interaction with immobilized DNA, the same amount of agent or NS DNA must be added to the experimental buffer to minimize refractive index differences.
2. The sample concentration depends on the magnitude of the binding constant (K_A). With a single binding site, for example, concentrations at least ten times above and below $1/K_A$ should be used (i.e., a 100-fold difference between the lowest and highest concentrations). A larger concentration range above and below $1/K_A$ will yield a more complete binding curve. For binding

Table 2
Regeneration solutions

Interaction strength	Acidic	Basic	Hydrophobic	Ionic
Weak	pH > 2.5	pH < 9	pH < 9	
	10 mM glycine/HCl HCl Formic acid	10 mM HEPES/NaOH	50 % ethylene glycol	1 M NaCl
Intermediate	pH 2–2.5	pH 9–10	pH 9–10	
	10 mM glycine/HCl Formic acid HCl H_3PO_4	10 mM glycine/NaOH NaOH	50 % ethylene glycol	2 M $MgCl_2$
Strong	pH < 2 10 mM glycine/HCl HCl Formic acid H_3PO_4	pH > 10 NaOH	pH > 10 25–50 % ethylene glycol	4 M $MgCl_2$ 6 M guanidine-hydrochloride

constants of 10^7–10^9 M^{-1}, as observed with many nucleic acid/protein complexes, protein concentrations from 0.01 nM to 10 μM in the flow solution allow accurate determination of binding constants (*see* Subheading 1.3.1). Injecting samples from low to high concentration is useful for eliminating artifacts in the data from adsorption or carry over (*see* **Note 8**).

3. Possible problems at high sample concentrations: poor sensorgrams and nonspecific binding may be obtained. For proteins that self-associate, it is important to maintain concentrations well below levels at which oligomerization occurs. For example, the PU.1 ETS domain is known to dimerize at above 10 μM and above 1 μM in the DNA bound-state [29]. Therefore, the sample concentrations of PU.1 in this protocol are maintained well below this level (<0.4 μM) to ensure that the protein presents as a monomer in both free and bound states.

3.4 Regeneration

1. Regeneration is the process of removing bound analyte from the sensor chip surface after analysis of a sample, in preparation for the next analysis cycle.
2. Regeneration conditions should remove the bound analyte completely from the surface without destroying the immobilized reagent. Generally used regeneration solutions are listed in Table 2. In general, milder conditions are initially used, and more harsh conditions are applied as needed. Some other regeneration solutions for special samples are available from

the Biacore website. In our studies, 1 M NaCl solution is typically used as a gentle but efficient regeneration solution to remove protein from the DNA immobilized sensor chip surface (*see* **Note 5**).

3. Injections of 30–60 s of regeneration solution are usually sufficient. Longer exposure to regeneration conditions involves greater risks of losing binding capacity on the surface, and often does not improve regeneration.
4. After injection of regeneration solution, three 1-min injections of experimental buffer are recommended to reducing any remaining regeneration solution.
5. At the end of each cycle, 5 min running with buffer flowing ensures that the chip surface is re-equilibrated for binding (i.e., the dextran matrix is re-equilibrated with experimental buffer) and the baseline has stabilized before the next sample injection.

3.5 Data Collection and Processing

The Biacore control software allows users to write a method or to use a method wizard to set up experiments. Several key factors, such as flow rate, flow path, association and dissociation time, injection order, surface regeneration and post-regeneration re-equilibration, must be considered in setting up experiments. An example of the method used to collect PU.1 binding data on DNA surface is shown below.

1. A Biacore T200 instrument (GE Healthcare Inc.) is used as an example in this protocol. This is one of the most sensitive biosensor-SPR instruments but other instruments are excellent and the application will determine what sensitivity is needed. Binding small molecules to a protein or DNA requires high sensitivity while the interaction of two proteins, a protein with DNA or other macromolecule complexes, requires less sensitivity.
2. Streptavidin is immobilized on a CM4 sensor chip as described in Subheading 3.1, and then three biotin-labeled hairpin DNAs are immobilized in different flow cells as described in Subheading 3.2. The biotin-labeled target DNA sequence discussed in this example is shown in Fig. 2. This DNA is based on the λB motif of the lg2-4 enhancer [34], 5′-AAAGGAAGTG-3′, a native high-affinity cognate site for PU.1 [35, 36].
3. 1 M NaCl is used as regeneration solution.
4. The ETS domain of PU.1 (residues 167–262 from the murine sequence) is overexpressed in *Escherichia coli* and purified as previously described [28].
5. Na_2HPO_4 buffer (25 mM Na_2HPO_4, 1 mM EDTA and 0.05 % (v/v) surfactant P20, pH 7.4) with different concentrations of

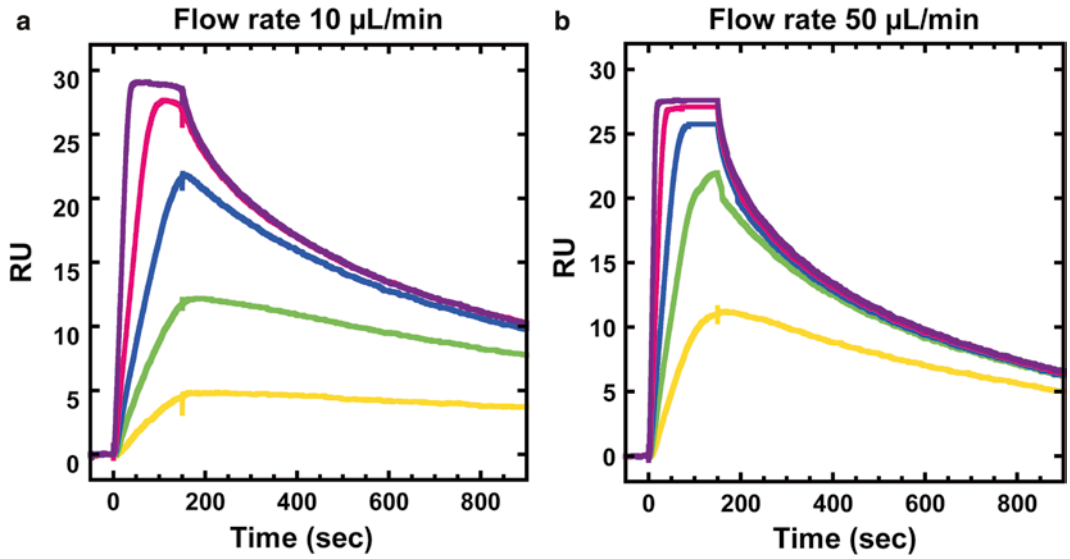


Fig. 3 Sensorgrams of PU.1 binding with target DNA sequence with 300 mM NaCl at flow rate of (**a**) 10 µL/min and (**b**) 50 µL/min. The PU.1 concentrations from bottom to top are 1, 2, 3, 5, and 10 nM in both plots

NaCl have been used as experimental and experimental buffer to optimize ionic conditions. An example of mass transport effects on PU.1-DNA interaction with 300 mM NaCl is shown in Fig. 3. It is clear that the shapes of sensorgram and apparent kinetics are strongly dependent on flow rates (*see* Subheading 1.3.2). Sensorgrams with the same samples at other flow rates, such as 75 and 100 µL/min, have also been run with 300 and 400 mM NaCl and they are very similar (data not shown), suggesting that by increasing the flow rate the mass transport effects could be minimized and finally removed. Therefore, a high flow rate (100 µL/min) is employed in this protocol and Na_2HPO_4 buffer containing 400 mM NaCl, with and without 300 µM/base pair of salmon sperm DNA as nonspecific DNA, is used as the experimental buffer (*see* **Note 9**).

6. A series of concentrations (concentration range is from 1 nM to 400 nM) of PU.1 is prepared with the experimental buffer to cover the concentration range around the K_D at any salt concentration (*see* Subheading 1.3.1).
7. The flow rate is set to 100 µL/min to minimize mass transport effects.
8. Several buffer samples are injected at the start of each experiment as a baseline stabilization step. At the beginning of each sample injection cycle, experimental buffer flows over the sensor chip surface for 5 min to give a very stable baseline that is essential for accurate binding analysis.

9. Inject 250 μL of each PU.1 solution at each concentration and set 600 s as the dissociation time (*see* **Note 10**). Protein samples are injected from low to high concentration to eliminate artifacts in the data from adsorption carry-over on the instrument flow system.
10. Inject regeneration solution (1 M NaCl) for 1 min in the end of the dissociation phase, followed by three 1-min experimental buffer injections to produce a stable baseline for the next sample cycle.
11. After the data are collected, open the experimental sensorgrams in the Biacore evaluation software for data processing (*see* **Note 11**). Zero the baselines on the response (*y*-) and the time (*x*-) axes by choosing a small region of a few seconds for averaging prior to sample injection on both the sample and control flow cells.
12. Subtract the response of the reference flow cell (FC1) from the reaction flow cell (i.e., FC2-1, FC3-1, and FC4-1). This can remove the effects from any bulk shift contribution on the changes of RUs.
13. Subtract a buffer injection, or an average of several buffer injections from a series of ligand injections at different concentrations on the same reaction flow cell. The reference correction and the buffer correction are known as double subtraction and can eliminate specific baseline irregularities [8, 37]. At this stage, the data are ready for analysis as discussed below.

3.6 Data Analysis

1. After the sensorgrams are processed as described above, kinetic and/or steady state analysis is performed. Both kinetic and steady state fitting can be done in Biacore evaluation software or with the Scrubber-2 package written by Myszka and collaborators (http://www.biologic.com.au). As can be seen in Fig. 4a, b, PU.1 binding reaches a steady state plateau in the injection period so that both kinetic and steady state analyses can be used. In this case the binding rate is not limited by mass transfer and the association and dissociation rate constants can be determined. The average of the data over a selected time period in the steady state region of each sensorgram can be obtained, converted to r ($r = RU_{obs}/RU_{max}$, Eq. 5) and plotted as a function of protein concentration in the flow solution, as shown in Fig. 4c, d (*see* **Note 12**).
2. Equilibrium constants can be obtained by fitting the sensorgrams to the equivalent site model in Eq. 5.
3. The RU on the surface is directly indicating the amount of PU.1 bound. Based on the RU at saturation we can determine that PU.1 forms a 1:1 complex with target DNA as expected.
4. A global kinetic fit in a 1:1 model with mass transport is applied for the sensorgrams in Fig. 4a, b to determine the binding

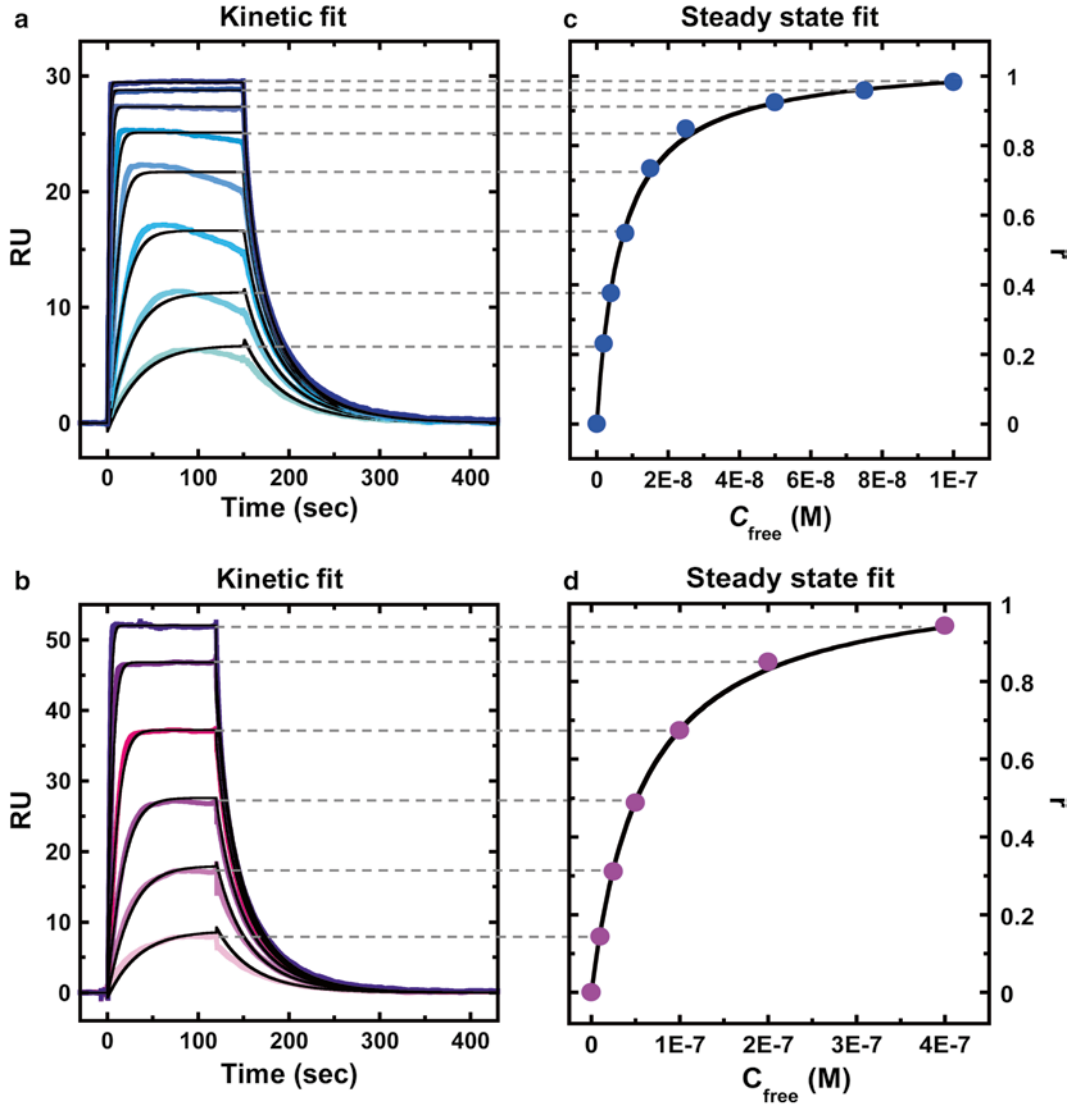


Fig. 4 Sensorgrams (color) and global kinetic fitting (black overlays) in a 1:1 binding model for PU.1 binding to the target DNA sequence with (**a**) 400 mM NaCl and (**b**) 400 mM NaCl in the presence of 300 μM bp salmon sperm DNA as nonspecific DNA**.** The PU.1 concentrations from bottom to top are (**a**) 2, 4, 8, 15, 25, 50, 75, and 100 nM, and (**b**) 10, 25, 50, 100, 200, and 400 nM. (**c** and **d**) Steady state fitting of the sensorgrams in **a** and **b**, respectively. RU values from the steady state region were converted to ***r*** (***r*** = RU_{obs}/RU_{max}) and are plotted as a function of unbound protein concentration with equilibrium in the complex

kinetics and affinity, and the results are listed in Table 3. With 400 mM NaCl, PU.1-DNA interaction has a very fast association [$k_a = (3.1 \pm 0.1) \times 10^7$ M^{-1} s^{-1}] and an overall binding affinity of 7.0 nM. In the presence of nonspecific DNA, the dissociation of PU.1 is barely affected, while the association rate is decreased around tenfold resulting in a 10-time weaker

Table 3
Binding affinities and kinetics for PU.1 with and without nonspecific (NS) DNA at 400 mM NaCl[a]

			K_D (nM)	
	K_a ($\times 10^6$ M^{-1} s^{-1})	k_d (s^{-1})	Kinetic fit	Steady-state fit
No NS DNA	31 ± 1	0.22 ± 0.01	7.0 ± 0.2	6.8 ± 0.3
300 μM NS DNA	2.4 ± 0.1	0.18 ± 0.01	73 ± 1	63 ± 2

[a]Errors listed in this table are standard errors for the fit of 1:1 binding model

binding affinity ($K_D = 73 \pm 1$ nM) compared to PU.1 binding without nonspecific DNA. The steady state plateau is obtained for every injection in Fig. 4a, b, which allows the steady state fits to be performed (Fig. 4c, d). The binding affinity values are in excellent agreement to the results determined by kinetic rate constants ($K_D = k_d/k_a$, Table 3). This suggests that the mass transfer effect is not significant and does not dominate the kinetics evaluation in these experiments (*see* **Notes 13** and **14**).

5. Comments on more complex binding models can be input through Biacore or other evaluation software.

4 Notes

1. Maintenance chips are available from GE Healthcare Inc. "Desorb" is a Biacore software command that instructs the instrument to remove adsorbed proteins from the flow system. A detailed list of commands and operations are shown in Table 1. Make sure that the analysis and sample compartment temperatures are not below 20 °C, since SDS in Desorb solution 1 may precipitate at low temperature. For biological samples such as protein, the Sanitize method should also be used after Desorb to insure that there is no protein left for microorganisms to grow in the liquid injection and flow system.
2. After running the regular Desorb, if the baseline is still not stable with ±1.0 RU/min, an additional super clean method may be used. First, run Desorb using 1 % (v/v) acetic acid in place of desorb solutions 1 and 2 with deionized water at 50 °C followed by one Prime to wash out the residual acetic acid. Then, run Desorb using 0.2 M sodium bicarbonate followed by one Prime to wash out the residual sodium bicarbonate. Finally, run Desorb using 6 M guanidine HCl in place of SDS (solution 1) and 10 mM HCl for glycine (solution 2). Prime the instrument a few times to thoroughly clean all residue and flow buffer to stabilize the instrument.

3. The choice of sensor chip depends on the nature and demands on the application. For general purposes, a Biacore CM5 sensor chip, which carries a matrix of carboxymethylated (CM) dextran covalently attached to the gold surface, can be used. It has a high surface capacity for immobilizing a wide range of ligands from protein to nucleic acids and carbohydrates. For protein–DNA interaction investigation, the Biacore CM4 sensor chip is another good choice because it is similar to sensor chip CM5 but has a lower degree of carboxymethylation (~30 % of that of CM5 chip) and charge that helps reduce nonspecific binding of the highly positively charged molecules, such as protein, to the surface. Sensor chip SA has a surface carrying a dextran matrix to which streptavidin has been covalently attached. Streptavidin has a very high binding affinity with biotin ($K_D \approx 10^{-15}$ M), so that the surface provides a high capture of biotinylated ligands. The SA chip is particularly suited to work with nucleic acids since 5′ or 3′ terminal biotinylation of nucleic acid is a well-established procedure.
4. The selection of experimental buffer depends on the nature of the target protein and DNA sequence. Salt concentration can be adjusted based on experimental requirements. With increasing of salt concentration, the binding affinity for protein–DNA interaction typically decreases for positively charged protein.
5. Regeneration conditions must be harsh enough to disrupt the complex and remove the bound reagent but mild enough to keep the DNA strand intact. It is highly recommended to start with the mildest conditions and short surface contact times since regeneration solutions can degrade DNA or immobilized matrix.
6. The same procedure can be applied to immobilize other protein or DNA that is labeled with a terminal amino group. Other molecules with a free amino group can also be captured by this method.
7. The Biacore T200 is good for experiments with less than 10 RU, while the Biacore 3000 and X100 have around 1/10th sensitivity of Biacore T200. Another way to calculate the amount of DNA immobilized is by using a standard ligand with known binding stoichiometry and binding affinity such as DNA minor groove binder netropsin to titrate the amount of DNA on the surface [31]. This is especially useful as sensor chips are re-used and begin to lose immobilized DNA.
8. To estimate an unknown K_A, it is necessary to conduct a preliminary experiment with several samples in concentrations spanning a broad range. Subsequently, a more focused set of concentrations covering the binding range is run to determine the K_A accurately.

9. As mentioned above (*see* **Note 4**), the salt concentration in buffer can be adjusted for experimental requirements. Here, the strong binding of PU.1 with its native high-affinity DNA needs at least 400 mM NaCl to be evaluated by SPR without mass transport effects. By conducting the experiments at different salt concentrations, a linear log K_D vs. log [Na+] plot is obtained that can be extrapolated to lower salt concentrations [27].
10. A sufficient association phase with a plateau region is needed for steady state analysis. For the most accurate fitting of the dissociation phase it is highly recommended to allow sufficient time for the protein to dissociate at least 50 % from the complex.
11. Other software programs such as Scrubber 2 and CLAMP are available for processing Biacore data. The results can also be exported and presented in graphing software such as KaleidaGraph. For the Biacore T200 user, data processing can be automatically performed with the Biacore T200 evaluation software, which is sufficient for most routine analyses and much more convenient for new users.
12. In some instances at low concentrations where the response does not reach steady state, the equilibrium responses can be estimated from kinetic fits of the sensorgrams using the known RU_{max} from the higher concentration sensorgrams. This extrapolation method works well with sensorgrams where the observed response is at least 50 % of the equilibrium RU.
13. The different RU_{max} values at PU.1 binding saturation in Fig. 4a, b are due to the different amounts of immobilized DNA left on the experimental sensor chips. As mentioned in **Note 7**, through appropriate calibration of the amount of immobilized DNA on the flow cell, binding kinetics and affinity analysis can be accurately performed.
14. Accurate sensorgrams and binding curves can be obtained at several salt concentrations and extrapolated to any other salt concentrations using standard methods [27]. *See* **Note 9**.

Acknowledgments

We gratefully thank the NIH (GM111749) and the NSF (MCB1411502) for the support for biosensor-SPR studies on DNA complexes, and the Georgia Research Alliance for funding of the Biacore instruments. We thank Carol Wilson for manuscript proofreading.

References

1. Homola J (2008) Surface plasmon resonance sensors for detection of chemical and biological species. Chem Rev 108:462–493
2. Rich RL, Myszka DG (2000) Advances in surface plasmon resonance biosensor analysis. Curr Opin Biotechnol 11:54–61
3. Wilson WD (2002) Analyzing biomolecular interactions. Science 295:2103–2105
4. Piliarik M, Vaisocherova H, Homola J (2009) Surface plasmon resonance biosensing. In: Rasooly A, Herold KE (eds) Methods in molecular biology, vol 503. Humana, New York, pp 65–88
5. Davis TM, Wilson WD (2001) Surface plasmon resonance biosensor analysis of RNA-small molecule interactions. Methods Enzymol 340:22–51
6. Papalia GA, Giannetti AM, Arora N, Myszka DG (2008) Thermodynamic characterization of pyrazole and azaindole derivatives binding to p38 mitogen-activated protein kinase using Biacore T100 technology and van't Hoff analysis. Anal Biochem 383:255–264
7. Liu Y, Wilson WD (2010) Quantitative analysis of small molecule-nucleic acid interactions with a biosensor surface and surface plasmon resonance detection. In: Fox KR (ed) Methods in molecular biology, vol 613. Humana, New York, pp 1–23
8. Nanjunda R, Munde M, Liu Y, Wilson WD (2011) Real-time monitoring of nucleic acid interactions with biosensor-surface plasmon resonance. In: Wanunu M, Tor Y (eds) Methods for studying nucleic acid/drug interactions. CRC, Boca Raton, pp 91–122
9. He X, Coombs D, Myszka DG, Goldstein B (2006) A theoretical and experimental study of competition between solution and surface receptors for ligand in a Biacore flow cell. Bull Math Biol 68:1125–1150
10. Bondeson K, Frostellkarlsson A, Fagerstam L, Magnusson G (1993) Lactose repressor-operator DNA interactions: kinetic analysis by a surface plasmon resonance biosensor. Anal Biochem 214:245–251
11. Goeddel DV, Yansura DG, Caruthers MH (1977) Binding of synthetic lactose operator DNAs to lactose represessors. Proc Natl Acad Sci U S A 74:3292–3296
12. Davis TM, Wilson WD (2000) Determination of the refractive index increments of small molecules for correction of surface plasmon resonance data. Anal Biochem 284:348–353
13. Degnan BM, Degnan SM, Naganuma T, Morse DE (1993) The ETS multigene family is conserved throughout the metazoa. Nucleic Acids Res 21:3479–3484
14. Oikawa T, Yamada T (2003) Molecular biology of the Ets family of transcription factors. Gene 303:11–34
15. Sharrocks AD (2001) The ETS-domain transcription factor family. Nat Rev Mol Cell Biol 2:827–837
16. Sementchenko VI, Watson DK (2000) Ets target genes: past, present and future. Oncogene 19:6533–6548
17. Hsu T, Trojanowska M, Watson DK (2004) Ets proteins in biological control and cancer. J Cell Biochem 91:896–903
18. Gilliland DG (2001) The diverse role of the ETS family of transcription factors in cancer. Clin Cancer Res 7:451–453
19. Oikawa T (2004) ETS transcription factors: possible targets for cancer therapy. Cancer Sci 95:626–633
20. Galang CK, Muller WJ, Foos G, Oshima RG, Hauser CA (2004) Changes in the expression of many Ets family transcription factors and of potential target genes in normal mammary tissue and tumors. J Biol Chem 279: 11281–11292
21. Wei G-H, Badis G, Berger MF, Kivioja T, Palin K, Enge M et al (2010) Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J 29:2147–2160
22. Hollenhorst PC, McIntosh LP, Graves BJ (2011) Genomic and biochemical insights into the specificity of ETS transcription factors. In: Kornberg RD, Raetz CRH, Rothman JE, Thorner JW (eds). Annu Rev Biochem 80:437–471
23. DeKoter RP, Singh H (2000) Regulation of B lymphocyte and macrophage development by graded expression of PU.1. Science 288: 1439–1441
24. Ross IL, Yue X, Ostrowski MC, Hume DA (1998) Interaction between PU.1 and another Ets family transcription factor promotes macrophage-specific basal transcription initiation. J Biol Chem 273:6662–6669
25. Kopp JL, Wilder PJ, Desler M, Kim JH, Hou J, Nowling T et al (2004) Unique and selective effects of five Ets family members, Elf3, Ets1, Ets2, PEA3, and PU.1, on the promoter of the type II transforming growth factor-beta receptor gene. J Biol Chem 279:19407–19420
26. Pham TH, Minderjahn J, Schmidl C, Hoffmeister H, Schmidhofer S, Chen W et al (2013) Mechanisms of in vivo binding site selection of the hematopoietic master

transcription factor PU.1. Nucleic Acids Res 41:6391–6402

27. Munde M, Poon GM, Wilson WD (2013) Probing the electrostatics and pharmacological modulation of sequence-specific binding by the DNA-binding domain of the ETS family transcription factor PU.1: a binding affinity and kinetics investigation. J Mol Biol 425:1655–1669
28. Poon GM (2012) Sequence discrimination by DNA-binding domain of ETS family transcription factor PU.1 is linked to specific hydration of protein-DNA interface. J Biol Chem 287:18297–18307
29. Poon GM (2012) DNA binding regulates the self-association of the ETS domain of PU.1 in a sequence-dependent manner. Biochemistry 51:4096–4107
30. Myszka DG (2000) Kinetic, equilibrium, and thermodynamic analysis of macromolecular interactions with BIACORE. Methods Enzymol 323:325–340
31. Nguyen B, Tanious FA, Wilson WD (2007) Biosensor-surface plasmon resonance: quantitative analysis of small molecule-nucleic acid interactions. Methods 42:150–161
32. Tanious FA, Nguyen B, Wilson WD (2008) Biosensor-surface plasmon resonance methods for quantitative analysis of biomolecular interactions. In: Correia JJ, Detrich HW (eds). Methods Cell Biol 84:53–77
33. Karlsson R (1999) Affinity analysis of non-steady-state data obtained under mass transport limited conditions using BIAcore technology. J Mol Recognit 12:285–292
34. Eisenbeis CF, Singh H, Storb U (1993) PU-1 is a component of a multiprotein complex which binds an essential site in the murine immunoglobulin lambda-2-4 enhancer. Mol Cell Biol 13:6452–6461
35. Poon GM, Gross P, Macgregor RB (2002) The sequence-specific association of the ETS domain of murine PU.1 with DNA exhibits unusual energetics. Biochemistry 41: 2361–2371
36. Poon GM, Macgregor RB (2003) Base coupling in sequence-specific site recognition by the ETS domain of murine PU.1. J Mol Biol 328:805–819
37. Myszyka DG (1999) Improving biosensor analysis. J Mol Recognit 12:279–284

Chapter 21

Identification of Nucleic Acid High Affinity Binding Sequences of Proteins by SELEX

Philippe Bouvet

Abstract

A technique is described for the identification of nucleic acid sequences bound with high affinity by proteins or by other molecules suitable for a partitioning assay. Here, a histidine-tagged protein is allowed to interact with a pool of nucleic acids and the protein–nucleic acid complexes formed are retained on a Ni-NTA matrix. Nucleic acids with a low level of recognition by the protein are washed away. The pool of recovered nucleic acids is amplified by the polymerase chain reaction and is submitted to further rounds of selection. Each round of selection increases the proportion of sequences that are avidly bound by the protein of interest. The cloning and sequencing of these sequences finally completes their identification.

Key words SELEX, Binding sequences, Sequence recognition, Nucleic acids ligands

1 Introduction

The interactions of nucleic acids with proteins are involved in numerous biological functions. Most of these interactions involve specific contacts between nucleic acid and protein with variable binding affinities. The identification of nucleic acid recognition sequence of a specific protein is often the first step to undertake the study of the biological function of this protein. Over the last 15 years, the SELEX procedure (*S*ystematic *E*volution of *Li*gands by *EX*ponential enrichment) has been used to identify high affinity nucleic acids ligands for different proteins. This method was first described for the selection of DNA and RNA target of nucleic acid binding proteins [1, 2] but has then been used for the selection of nucleic acid ligands for any kind of targets [3]. This methodology uses the power of genetic selection techniques combined with the advantage of in vitro biochemical experiments. It is a rapid in vitro technique, and is relatively easy to implement in all laboratories. This procedure should accelerate and simplify nucleic acid–protein interaction studies.

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4_21,

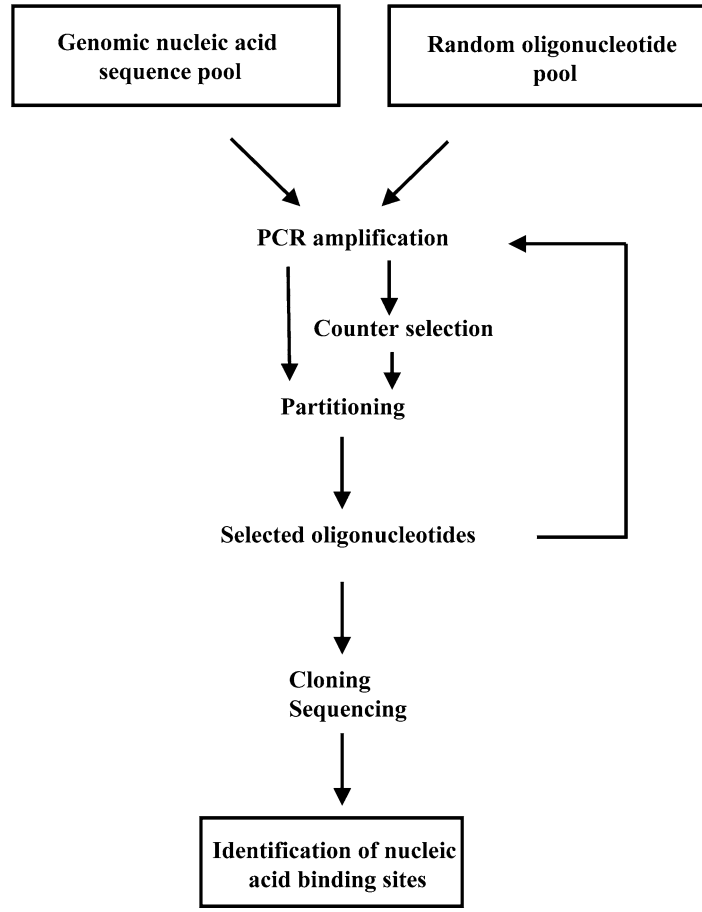


Fig. 1 Schematic representation of the different steps of SELEX

This process involves few simple steps as described in Fig. 1. The procedure consists of an enrichment of individual oligonucleotide molecules from complex mixtures of nucleic acid sequences by repeated rounds of selection. First, ligand sequences that bind to the target protein are selected. Then the bound and the unbound molecules are partitioned. In a final step, the selected sequences are amplified by PCR. This cycle of selection-amplification (round) is repeated until enrichment is obtained for nucleic acid sequences that bind to the protein with high affinity.

The strategy is designed to determine the optimal binding nucleic acid sequences, also called "aptamer" [4]. However, it should be noted, that high affinity nucleic acid ligand could be isolated even for partitioning agents which are not nucleic acid binding proteins, and the isolated SELEX sequences may not be related to the real binding site of the protein in vivo. Therefore, if this strategy is used to determine the nucleic acid recognition sequence of a protein, one must realize that the most difficult part

of this study will be the analysis of the isolated SELEX sequences and the demonstration that there are relevant for the in vivo function of the studied protein.

Numerous protocols for the SELEX procedure have been used successfully by different laboratories. In fact, each step of this procedure can be optimized [5] and modified in function of the characteristics of the nucleic acid binding protein that is studied [6]. In the initial SELEX protocol, the aptamer libraries were made with chemically synthesized nucleic acid molecules. In genomic SELEX, genome derived nucleic acid sequences (DNA or RNA) that bind with affinity to specific protein are identified [7, 8]. More recently, selection procedures using microfluidic devices such as capillary chromatography and microarrays have also been developed [9]. Several automated procedures have also been proposed [10–12] but these technologies cannot yet be routinely used in any molecular biology laboratory.

We will provide here a basic typical protocol that has been used successfully by several laboratories [13–17] to identify nucleic acid ligands for RNA and DNA binding proteins and that can be very easily implemented in any laboratory to perform the selection procedure in just a few days. Typically, one round of selection can be done in one day. In this protocol, we propose to use an oligonucleotide template that contains a random sequence of 25 nucleotides. This oligonucleotide has been used in many studies for the identification of RNA and DNA binding proteins [13–15, 18] but it can be replaced by any other DNA template with variable random nucleotide length that would be more suitable for the protein that is being studied. We are providing a detailed protocol for the selection of aptamers of RNA binding proteins. In the case of DNA binding proteins, steps 6 and 18 of the method should be omitted. For each step, detailed descriptions of the strategic choices that should be made and of modifications that can be introduced in the protocol will be given in Subheading 4.

2 Materials

1. The following synthetic DNA template has been used with success by several laboratories:

   5′ TGGGCACTATTTATATCAAC (N25) AATGTCGTTG GTGGCCC 3

   with these flanking primers:

   T7 5′-CGCGGATCCTAATACGACTCACTATAGGGGCC ACCAACGACATT-3′

   and Rev 5′-CCCGACACCCGCGGATCCATGGGCACTA TTTATATCAAC-3′.

The T7-Xba primer 5′ GGTCTAGATAATACGACTCA CTATAGGGG 3′ and

Rev-HIII primer 5′ ACCGCAAGCTTATGGGCACTAT TTATAT 3′ can be used for the final PCR amplification, and will allow an oriented cloning (*Xba*I and *Hin*dIII) of the PCR product in a cloning vector like pBluescript (Stratagene).

2. Thermocycler.
3. *Taq* polymerase.
4. Partitioning matrix (to be chosen in function of the studied protein).
5. Nucleic acid electrophoresis system.
6. NT2 buffer: 50 mM Tris–HCl (pH 7.4), 150 mM NaCl, 0.05 % NP 40, 1 mM $MgCl_2$.
7. Binding buffer (BB): 50 mM Tris–HCl (pH 7.5), 150 mM NaCl, 20 mM KCl, 1 mM DTT, 0.05 % NP 40, 1 mM $MgCl_2$, 2.5 % polyvinyl alcohol (PVA), 1 mM EGTA, 50 µg/mL poly(A), 2 µL/mL vanadyl ribonucleoside complex (VRC), 0.5 µg/mL tRNA, 125 µg/mL BSA.
8. 5× reverse transcription buffer: 250 mM Tris–HCl (pH 8.5), 40 mM $MgCl_2$, 5 mM DTT, 250 µg/mL BSA, 150 mM KCl.
9. 1× Transcription buffer: 40 mM Tris–HCl (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mM NaCl, 10 mM DTT.
10. Gel shift buffer (GSB): 100 mM Tris–HCl (pH 7.4), 4 mM $MgCl_2$, 200 mM KCl, 20 % glycerol, 1 mM dithiothreitol, 0.5 mg/mL tRNA, 4 µg/mL BSA.
11. Polyacrylamide gel shift: 8 % polyacrylamide (acrylamide–bis, 60:1) containing 5 % glycerol in 0.5× TBE buffer (0.045 M Tris–borate, 1 mM EDTA).

3 Method

1. About 10 pmol of synthetic template DNA (N25) (*see* **Notes 1 and 2**) are amplified by PCR in a 100 µL reaction in a 500 µL test tube (*see* **Note 3**). Add 2 µL of a mix of all 4 four dNTPs (10 mM each) and 500 ng of each primer. 1 unit of *Taq* polymerase is added just before the start of the amplification procedure. If the thermocycler does not possess a hot cover, the reaction mixture is overlaid with two drops of mineral oil.
2. Set up the thermocycler with the following cycle conditions: denaturation 1 min at 94 °C, annealing 1 min at 50 °C, elongation 1 min at 72 °C for 25 cycles, then finish with an elongation of 10 min at 72 °C. At the end of the PCR reaction, the reaction can be left at 4 °C without further purification.

3. Analyze 5 µL of the PCR reaction on a 3 % agarose gel (made with 1× TAE). Run in parallel a commercial DNA ladder which gives characteristic bands around 100 pb. The PCR reaction should give a nice signal at 108 bp.
4. Add 100 µL of phenol–chloroform (1:1) to the PCR reaction and mix vigorously for 1 min. After a 5 min centrifugation at 21,000×*g*, the upper aqueous phase is extracted one more time with 1 volume of phenol–chloroform. 10 µL of 3 M sodium acetate (NaOAc) pH 5.0 and 300 µL of cold ethanol are added. Allow DNA precipitation for at least 15 min at –20 °C.
5. The PCR product is recovered by centrifugation (15 min at 21,000×*g*), washed with 70 % ethanol, dried and resuspended in 10 µL of sterile water.
6. In vitro transcription. This step should be omitted for a SELEX with a DNA binding protein. 1 µg of PCR product from **step 5** are incubated in 1× transcription buffer with 0.5 mM of each rNTP, 1 unit of RNasin, and 20 units of T_7 RNA polymerase. The reaction is allowed for 1 h at 37 °C. The DNA template is eliminated by addition of 1 unit of RNase-free DNase for 10 more minutes at 37 °C. After 2 phenol–chloroform extractions, the RNA is purified through a G50 column (to remove most of unincorporated nucleotides) then precipitated with 0.1 volume of 3 M NaAc pH 5 and 2 vol. 100 % EtOH for 15 min at –20 °C. The RNA is pelleted for 15 min at 21,000×*g*, washed with 70 % EtOH, dried and resuspended in 20 µL of RNase-free water. 1 µL of the transcription reaction is loaded on a 3 % agarose gel to check the quality of the RNA and the RNA concentration can then be determined through UV absorption at 260 nm.
7. Preparation of the partitioning matrix. The nature of this matrix will depend of the protein that is used (*see* **Note 4**). We will provide here a detailed protocol for a selection procedure with a histidine-tagged protein. Other strategies are mentioned in **Note 4**. Take 2 µL of Ni-NTA agarose beads (Qiagen), and wash twice with 500 µL of sterile water to remove all storage buffer (by a 15 s centrifugation in a benchtop centrifuge). Then wash the beads twice with 500 µL of NT2 buffer. During the last wash, the solution is divided in two test tubes (tube A and B). Centrifuge and eliminate 150 µL of supernatant.
8. Add about 1 pmol of purified histidine-tagged protein to tube A. Incubate for 30 min at 4 °C on a roller to allow binding of the protein on the Ni-NTA beads.
9. Centrifuge tube A for 15 s, remove supernatant, then wash the beads twice with 500 µL of NT2 buffer to remove all unbound protein.
10. Centrifuge tubes A and B, and remove most of the NT2 supernatant to leave about 10 µL of buffer above the beads. Ni-NTA beads must be visible at the bottom of the tube.

11. Add 100 µL of BB buffer in tube B.
12. Add 15 µg of nucleic acid from **step 5** (for a DNA binding protein) or from **step 6** (for an RNA-binding protein) in tube B (*see* **Note 6**). Incubate for 5 min at room temperature. Centrifuge tube B for 15 s. Remove and save supernatant. This is the counter selection (*see* **Note 7**).
13. Add supernatant from **step 12** to the tube that contains the protein (tube A) (*see* **Note 8**). Incubate for 5 min at room temperature. Centrifuge tube A for 15 s, then remove and discard supernatant.
14. Add 1 mL of NT2 buffer (*see* **Note 9**) in tubes A and B. Mix well by inverting the tubes three times. Centrifuge for 15 s to pellet the Ni-NTA beads. Remove the supernatant as much as possible.
15. Repeat this wash four more times. At the last wash, transfer the nucleic acid–protein complex in a new test tube (*see* **Note 10**).
16. After the last wash, leave 100 µL of NT2 buffer in each tube. Add 100 µL of sterile water and 200 µL of phenol–chloroform (1:1). Vortex for 30 s and spin for 5 min at full speed. Repeat this extraction one more time.
17. Recover the upper aqueous phase, and add 2 µL of 1 M $MgCl_2$, 20 µL of 3 M NaOAc pH 5 and 700 µL of 100 % EtOH. Precipitate for at least 30 min at –20 °C, spin for 30 min at 21,000×*g*. Wash the pellet with 70 % EtOH, dry and resuspend it in 13 µL of sterile water.
18. This step (reverse transcription) should be omitted if the SELEX is performed with a DNA oligonucleotide. To each tube A and B, add 100 ng of Rev primer (in 1 µL), 2 µL of a dNTP mix (each dNTP at 10 mM), 4 µL of 5× reverse transcription buffer, 30 units of RNasin (Promega), and 25 units of AMV reverse transcriptase (Boehringer Mannheim). The reaction is set up for 5 min at 55 °C then 1 h at 42 °C. 5 µL of this reverse transcription reaction is used directly, without further purification, for the next PCR amplification.
19. Add to the recovered nucleic acid of **step 17** or **18** the reagent necessary for the PCR reaction, as described in **step 1**. Include a control reaction without oligonucleotide template to make sure that the PCR reaction is performed in good conditions. No PCR product should be obtained with tube B (*see* **Note10**).
20. Purify the PCR product obtained with tube A as described in **steps 4** and **5** and repeat several rounds of selection.
21. After several rounds of selection (usually between 4 and 10) check for an enrichment of the oligonucleotide pool in high affinity ligand for the target protein (for example, *see* Fig. 2 and ref. 14) (also *see* **Note 11**).

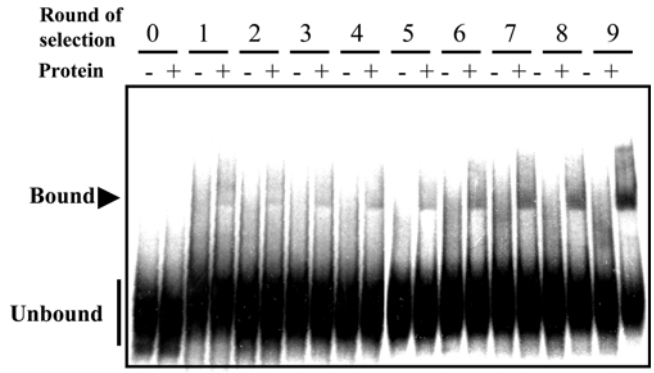


Fig. 2 Example of enrichment of the selected sequence after different rounds of amplification. An aliquot of each oligonucleotide pool after each round of amplification was used for in vitro transcription with [$\alpha-^{32}$P] CTP and used for gel shift assay. About 10 fmol of [^{32}P] RNA was incubated with (+) or without (–) protein (10 nM) for 15 min at room temperature in 20 μL of RNA binding buffer (GSB). Then the mixture was directly loaded on a 8 % polyacrylamide gel assay. This example corresponds to the selection that was performed with nucleolin [15]

22. The final PCR can be performed with the primers T7-Xba and Rev-HIII (*see* **Note 1**). After gel purification, the PCR products are digested with the restriction enzymes *Xba*I and *Hin*dIII, and the oligonucleotides are cloned in an adequate vector (for example pBluescript). Individual clones are selected and inserts are sequenced using standard methodology (*see* **Note 12**).

4 Notes

1. A custom made random oligonucleotide can be easily synthesized using standard chemistry. The variable sequence is flanked by fixed sequences at its 5′ and 3′ extremities. Full-length oligonucleotides should be gel purified before their amplification by PCR. The length of the random sequence may vary from a few nucleotides to as much as a hundred. If a simple DNA binding site is expected, a random sequence of as little as 20 nucleotides should be sufficient. A library with a short random sequence has also the advantage of being more likely to contain all possible random sequences and therefore to allow the selection of the best binding sequence. Sequencing of a few random sequences should be ideally performed to ensure that the synthesis of the random sequences has not been biased by a preferential incorporation of one deoxynucleotide. If the composition of the random sequence is severely biased, this should be corrected by modifying the percentage of addition of nucleotides accordingly during the synthesis of the random

sequence. If a binding site is already known for the protein, the SELEX procedure can be used to determine nucleotides important for binding affinity and specificity. In this case, an oligonucleotide containing a degenerated sequence within the known binding site can be synthesized [20]. Libraries that contain genomic sequences can also be used to identify potential natural binding sites [21, 22].

2. The random sequence is flanked by fixed sequences (17–20 nt) to allow PCR amplification with the corresponding primers (*see* **Note 1** for an example of primer sequences). It is important to check that the flanking sequences are not a binding site for the protein. This can easily be done by performing a binding assay between the random pool and the studied protein.
3. The PCR reaction can be modified to allow the production of single stranded oligonucleotides, the incorporation of modified nucleotides or random mutations, etc. If the PCR reaction produces aberrant products (higher molecular weight DNA products, smear, etc.) several tests reactions (with various amount of primers, number of cycle) can be realized.
4. Several methods of partitioning can be used in function of what is available to the researcher [19]. If the target is a tagged recombinant protein (GST, Histidine, or any other tag) nucleic acid–protein complexes can be recovered by classical affinity chromatography as described in this protocol. If the protein is pure but not tagged, filtration of the binding reaction mixture through nitrocellulose filter allows separation of the bound and unbound molecules [25]. An alternative method of partitioning uses gel shift analysis. For this method, labeled oligonucleotides are preferentially used to easily identify the nucleic acid–protein complex. Shifted oligonucleotides are eluted from the gel and used for PCR amplification.
5. The composition of the binding buffer should be adapted to the protein that is being studied. The addition of nucleic acid competitors like tRNA, or homopolymers like poly(A) or poly (dI-dC), might be necessary in some selection experiments to reduce the nonspecific binding of the protein to the random oligonucleotide. Preliminary tests of interaction of the random pool with the protein can be performed with nucleic acid competitors to determine the best selection conditions and the concentration of these competitors that need to be added if required.
6. The amount of oligonucleotide present in the binding reaction should be in large excess over the protein. This ensures an efficient competition between ligands for the protein. The ratio oligonucleotide/protein is often comprised within the range

of 10 and 1000. The volume of the binding reaction should also be determined in function of the diversity of the library. A large binding reaction volume might be required if one wants to test all possible sequences ($4n$, where n is the number of random nucleotides) present in the initial library.

7. The interaction of the random oligonucleotide pool with the partitioning matrix, without the protein (called counter selection or negative selection) (Fig. 1), is important to remove from the random oligonucleotide pool molecules with high affinity for the partitioning matrix. This counter selection is not necessary if the DNA bound to the protein is recovered using gel shift since only the shifted band will be used for the next cycles. The counter selection can be performed during the first rounds of selection and could be omitted for the next cycles.
8. In most published experiments, the SELEX procedure is performed with purified recombinant proteins. However, the nucleic acid binding specificity and affinity can sometimes be the result of interactions between the protein and other cellular polypeptides. The SELEX procedure can be performed with crude cell extracts that contain the protein of interest or multi-protein complexes if the partitioning procedure allows a specific recovering of the protein target [23, 24]. Epitope-tagged protein can be expressed in cells, or added to a cell extract and used for the SELEX. For some SELEX experiments it could be also interesting to use truncated protein with only the nucleic acid binding domain. In some case, this can significantly reduce nonspecific binding of the random oligonucleotide pool with the protein and therefore reduce the number of rounds necessary for the isolation of specific ligands.
9. The stringency of the binding and washing buffer can be increased if necessary. This could be done by increasing the salt concentration, or by adding 0.5–1.0 M urea. Usually, between 1 and 10 % of the initial oligonucleotide pool bind to the target protein. Preliminary binding tests should determine the optimal buffer stringency to allow a binding which falls within this range. Buffer stringency can also be increased during the cycling process if no substantial enrichment is observed.
10. It has been sometimes observed that nucleic acids bind poorly to some plastic tubes. This binding is however sufficient to increase the nonspecific binding and to give a PCR product after the amplification reaction. This problem could be overcome by using siliconized test tubes, or by transferring the nucleic acid–protein complex in a new test tube during the washing procedure.
11. After several rounds of selection (usually between 4 and 10) it is important to check for an enrichment of the oligonucleotide pool

in high affinity ligands for the target protein before proceeding to the cloning and sequencing steps of these selected sequences. This can be done by doing an interaction between the protein and labeled oligonucleotide pools of each round of selection (5′ labeling for DNA oligonucleotide or in vitro labeled transcription for RNA oligonucleotide). An example is shown in Fig. 2. This example corresponds to the selection that was performed with nucleolin [15]. An aliquot of each oligonucleotides pool after each round of amplification was used for in vitro transcription with [α−^{32}P] CTP and used for gel shift assay. About 10 fmol of [^{32}P] RNA were incubated with or without 10 nM of protein for 15 min at room temperature in 20 μL of RNA binding buffer (GSB, *see* Materials & Methods and **Note 10**). The mixture was then directly loaded on an 8 % polyacrylamide gel (acrylamide–bis, 60:1) containing 5 % glycerol in 0.5× TBE buffer. The gel was then dried and subjected to autoradiography. In this particular experiment, the selection procedure was stopped after nine rounds of selection when a significant enrichment of bound oligonucleotides was obtained. Subsequent sequencing of selected molecules showed that about 20 % of them contained a consensus motif. If no enrichment is observed, more rounds of selection can be performed in the same experimental conditions or with higher stringency (*see* **Note 7**).

12. The number of individual clones that need to be sequenced to identify a consensus binding site might vary from one experiment to the other. But, in general, if an enrichment of the selected sequences has been detected during the different rounds of selection (*see* **Note 11**) it might be possible to detect a consensus binding site in as few as 20 individual sequences. For the determination of a more precise consensus motif, more sequences will be needed. It might also happen that exactly the same selected sequence over the full length of the initial random region is found several times in the sequenced clones and it is therefore not possible to identify the consensus motif recognized by the protein since they have all the same sequence. It might indicate that the selection has been too strong, and in this case it will be required to perform the sequencing from the previous round of selection.

Acknowledgments

The work in the authors' laboratory is supported by grants from the CNRS, ANR N° BLAN07-2_190263 and Association pour la Recherche sur le Cancer (ARC).

References

1. Oliphant AR, Brandl CJ, Struhl K (1989) Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 protein. Mol Cell Biol 9:2944–2949
2. Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249:505–510
3. Gold L, Polisky B, Uhlenbeck O, Yarus M (1995) Diversity of oligonucleotide functions. Annu Rev Biochem 64:763–797
4. Ellington AD, Szostak JW (1990) In vitro selection of RNA molecules that bind specific ligands. Nature 346:818–822
5. Irvine D, Tuerk C, Gold L (1991) SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. J Mol Biol 222:739–761
6. Stoltenburg R, Reinemann C, Strehlitz B (2007) SELEX—a (r)evolutionary method to generate high-affinity nucleic acid ligands. Biomol Eng 24:381–403
7. Kim S, Shi H, Lee DK, Lis JT (2003) Specific SR protein-dependent splicing substrates identified through genomic SELEX. Nucleic Acids Res 31:1955–1961
8. Shtatland T, Gill SC, Javornik BE, Johansson HE, Singer BS, Uhlenbeck OC, Zichi DA, Gold L (2000) Interactions of Escherichia coli RNA with bacteriophage MS2 coat protein: genomic SELEX. Nucleic Acids Res 28, E93
9. Mosing RK, Bowser MT (2007) Microfluidic selection and applications of aptamers. J Sep Sci 30:1420–1426
10. Cox JC, Ellington AD (2001) Automated selection of anti-protein aptamers. Bioorg Med Chem 9:2525–2531
11. Cox JC, Hayhurst A, Hesselberth J, Bayer TS, Georgiou G, Ellington AD (2002) Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer. Nucleic Acids Res 30, e108
12. Hybarger G, Bynum J, Williams RF, Valdes JJ, Chambers JP (2006) A microfluidic SELEX prototype. Anal Bioanal Chem 384:191–198
13. Bouvet P, Matsumoto K, Wolffe AP (1995) Sequence-specific RNA recognition by the Xenopus Y-box proteins. An essential role for the cold shock domain. J Biol Chem 270: 28297–28303
14. Clouaire T, Roussigne M, Ecochard V, Mathe C, Amalric F, Girard JP (2005) The THAP domain of THAP1 is a large C2CH module with zinc-dependent sequence-specific DNA-binding activity. Proc Natl Acad Sci U S A 102:6907–6912
15. Ghisolfi-Nieto L, Joseph G, Puvion-Dutilleul F, Amalric F, Bouvet P (1996) Nucleolin is a sequence-specific RNA-binding protein: characterization of targets on pre-ribosomal RNA. J Mol Biol 260:34–53
16. Triqueneaux G, Velten M, Franzon P, Dautry F, Jacquemin-Sablon H (1999) RNA binding specificity of Unr, a protein with five cold shock domains. Nucleic Acids Res 27:1926–1934
17. Tsai DE, Harper DS, Keene JD (1991) U1-snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. Nucleic Acids Res 19:4931–4936
18. Harper DS, Fresco LD, Keene JD (1992) RNA binding specificity of a Drosophila snRNP protein that shares sequence homology with mammalian U1-A and U2-B″ proteins. Nucleic Acids Res 20:3645–3650
19. Gopinath SC (2007) Methods developed for SELEX. Anal Bioanal Chem 387:171–182
20. Bartel DP, Zapp ML, Green MR, Szostak JW (1991) HIV-1 Rev regulation involves recognition of non-Watson-Crick base pairs in viral RNA. Cell 67:529–536
21. Gao FB, Carson CC, Levine T, Keene JD (1994) Selection of a subset of mRNAs from combinatorial 3′ untranslated region libraries using neuronal RNA-binding protein Hel-N1. Proc Natl Acad Sci U S A 91:11207–11211
22. Singer BS, Shtatland T, Brown D, Gold L (1997) Libraries for genomic SELEX. Nucleic Acids Res 25:781–786
23. Pollock R, Treisman R (1990) A sensitive method for the determination of protein-DNA binding specificities. Nucleic Acids Res 18: 6197–6204
24. Ringquist S, Jones T, Snyder EE, Gibson T, Boni I, Gold L (1995) High-affinity RNA ligands to Escherichia coli ribosomes and ribosomal protein S1: comparison of natural and unnatural binding sites. Biochemistry 34: 3640–3648
25. Tuerk C, Eddy S, Parma D, Gold L (1990) Autogenous translational operator recognized by bacteriophage T4 DNA polymerase. J Mol Biol 213:749–761

INDEX

Benoît P. Leblanc and Sébastien Rodrigue (eds.), *DNA-Protein Interactions: Principles and Protocols*, Methods in Molecular Biology, vol. 1334, DOI 10.1007/978-1-4939-2877-4, © Springer Science+Business Media New York 2015

G

H

I

K

L

M

N

P

Q

R

S

T

U

W

Y